

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: August 13, 1999, 21:54:13 ; Search time 909.15 Seconds  
(without alignments)  
12229.403 Million cell updates/sec  
Title: US-09-042-460-1  
Sequence: 1 GAATTCGGGTGGAGGCC.....CCGAGCTCGGTACCAAGCTT 3496  
Scoring table: IDENTITY\_NUC  
Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.\*

- 1: gb\_bal.\*
- 2: gb\_ba2.\*
- 3: gb\_om.\*
- 4: gb\_ov.\*
- 5: gb\_pat.\*
- 6: gb\_ph.\*
- 7: gb\_pl1.\*
- 8: gb\_pl2.\*
- 9: gb\_pr1.\*
- 10: gb\_pr2.\*
- 11: gb\_pr3.\*
- 12: gb\_ro.\*
- 13: gb\_st.\*
- gb\_sy.\*
- gb\_un.\*
- gb\_vl.\*
- fun.\*
- htg.\*
- umi.\*
- w2.\*

Pred. No. is the number of  
score greater than or equal to  
and is derived by analysis of the  
Result No. 1 3426  
Score Query Match Length DB ID  
3426 98.0 3426 12 AF051911  
SUMMARY chance to have a  
result being printed,  
distribution.

Description  
Mus muscu

us-09-042-460-1.rge

K AU 9 H A L  
09/04/1998  
SEQ ID: 1, 2

2	3367.4	96.3	3369	12	AF073311	AF073311 Mus muscu
3	1603.6	45.9	8960	15	AF043739	AF043739 Synthetic
4	1600.8	45.8	4015	11	AF015950	AF015950 Homo sapi
5	1462.6	41.8	4027	11	AF018167	AF018167 Homo sapi
6	468.8	13.4	7498	12	AF121949	AF121949 Homo sapi
7	468.8	13.4	12213	42	AF114847	AF114847 Homo sapi
8	468.8	13.4	12532	42	AF121948	AF121948 Homo sapi
9	437.4	12.5	26414	42	HSPERT1	AF128893 Homo sapi
10	194.6	5.6	5491	9	AB016767	AB016767 Homo sapi
11	154.8	4.4	201	12	AF029235	AF029235 Mus muscu
12	154.8	4.4	4356	11	AF097365	AF097365 Homo sapi
13	154.8	4.4	2043	11	AF098956	AF098956 Homo sapi
14	122.2	3.5	25138	42	HSPERT2	AF128894 Homo sapi
15	65.6	1.8	3418	10	HSP56NKR	AJ000542 Homo sapi
16	63	1.8	1102	8	AF137070	AF137070 Sphenosty
17	61.6	1.8	758	36	AAU60877	U60877 Anopheles a
18	59.6	1.7	339	2	AF114621	AF114621 Unculture
19	59.4	1.7	779	36	AAU60880	U60880 Anopheles a
20	58.8	1.7	461	2	AF018066	AF018066 Unculture
21	58.6	1.7	424	12	AF033115	AF033115 Mus muscu
22	57.8	1.7	1242	36	SSU52949	AF018036 Spisula sol
23	57.8	1.7	1631	2	AF018036	AF018036 Unculture
24	57.8	1.7	689	2	AF018044	AF018044 Unculture
25	57.8	1.7	719	2	AF018052	AF018052 Unculture
26	57.8	1.7	719	2	AF018054	AF018054 Unculture
27	57.8	1.7	1199	17	DVU93295	U93295 Dengue viru
28	57.8	1.7	1197	17	DVU93301	U93301 Dengue viru
29	57.8	1.7	1200	17	DVU93302	U93302 Dengue viru
30	57.8	1.7	776	36	AAU60876	U60876 Anopheles a
31	57.6	1.6	226	12	MUSFRE	L39119 Mus muscu
32	57.4	1.6	448	36	SRMETPANS	AF001134 Strongylo
33	57.2	1.6	731	2	AF056365	AF056365 Unculture
34	57.2	1.6	1088	2	AF056365	AF056365 Unculture
35	57.2	1.6	230	10	HUMX0133D	L33810 Human (clon
36	57.2	1.6	3942	15	AF073995	AF073995 Vibrio ch
37	56.8	1.6	1528	36	BLA35433	AJ003433 Branchios
38	56.8	1.6	3497	10	BSA35433	AJ003168 Homo sapi
39	56.8	1.6	1673	12	D88984S01	D88984 Mus muscu
40	56.8	1.6	573	12	MUSDIP	D4443 Mouse mRNa
41	56.8	1.6	3110	17	HVISE364A	L22944 Human immun
42	55.8	1.6	1229	7	HBR010397	AJ010397 Hevea bra
43	55.4	1.6	1169	12	MMU89924	U89924 Mus muscu
44	55	1.6	5025	1	ECY10545	Y10545 Cloning vec
45	55	1.6	1081	2	AF056364	AF056364 Unculture

ALIGNMENTS

RESULT	1	AF051911	3426 bp	mus muscu	reverse transcriptase mRNA, complete cds.
LOCUS		AF051911			
DEFINITION		Mus musculus telomerase reverse transcriptase mRNA, complete cds.			
ACCESSION		AF051911			
NID		92005591			
VERSION		AF051911.1			
KEYWORDS		house mouse, Mus musculus, Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
SOURCE		Greenberg, R.A., Allsopp, R.C., Chin, L., Morin, G.B. and DePinho, R.A. Expression of mouse telomerase reverse transcriptase during development, differentiation, and proliferation (bases 1 to 3426)			
ORGANISM		Greenberg, R.A., Allsopp, R.C., Chin, L., Morin, G.B. and DePinho, R.A. Direct Submission Submitted (02-MAR-1998) Microbiology and Immunology, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461, USA			
REFERENCE		1 (bases 1 to 3426)			
AUTHORS		Greenberg, R.A., Allsopp, R.C., Chin, L., Morin, G.B. and DePinho, R.A.			
TITLE		Expression of mouse telomerase reverse transcriptase during development, differentiation, and proliferation (bases 1 to 3426)			
JOURNAL		Greenberg, R.A., Allsopp, R.C., Chin, L., Morin, G.B. and DePinho, R.A. Microbiology and Immunology, 1300 Morris Park Ave, Bronx, NY 10461, USA			
REFERENCE		2 (bases 1 to 3426)			
AUTHORS		Greenberg, R.A., Allsopp, R.C., Chin, L., Morin, G.B. and DePinho, R.A.			
TITLE		Expression of mouse telomerase reverse transcriptase during development, differentiation, and proliferation (bases 1 to 3426)			
JOURNAL		Greenberg, R.A., Allsopp, R.C., Chin, L., Morin, G.B. and DePinho, R.A. Microbiology and Immunology, 1300 Morris Park Ave, Bronx, NY 10461, USA			
FEATURES		Location/Qualifiers			

481	Db	ACCTGCTGGCACACTGTGCTCTTTATCTCTGGTGGCCCCCAGCTGTGCTCTACCAAGGTGT	540
482	Db		541
483	Db		542
484	Db		543
485	Db		544
486	Db		545
487	Db		546
488	Db		547
489	Db		548
490	Db		549
491	Db		550
492	Db		551
493	Db		552
494	Db		553
495	Db		554
496	Db		555
497	Db		556
498	Db		557
499	Db		558
500	Db		559
501	Db		560
502	Db		561
503	Db		562
504	Db		563
505	Db		564
506	Db		565
507	Db		566
508	Db		567
509	Db		568
510	Db		569
511	Db		570
512	Db		571
513	Db		572
514	Db		573
515	Db		574
516	Db		575
517	Db		576
518	Db		577
519	Db		578
520	Db		579
521	Db		580
522	Db		581
523	Db		582
524	Db		583
525	Db		584
526	Db		585
527	Db		586
528	Db		587
529	Db		588
530	Db		589
531	Db		590
532	Db		591
533	Db		592
534	Db		593
535	Db		594
536	Db		595
537	Db		596
538	Db		597
539	Db		598
540	Db		599
541	Db		600
542	Db		601
543	Db		602
544	Db		603
545	Db		604
546	Db		605
547	Db		606
548	Db		607
549	Db		608
550	Db		609
551	Db		610
552	Db		611
553	Db		612
554	Db		613
555	Db		614
556	Db		615
557	Db		616
558	Db		617
559	Db		618
560	Db		619
561	Db		620
562	Db		621
563	Db		622
564	Db		623
565	Db		624
566	Db		625
567	Db		626
568	Db		627
569	Db		628
570	Db		629
571	Db		630
572	Db		631
573	Db		632
574	Db		633
575	Db		634
576	Db		635
577	Db		636
578	Db		637
579	Db		638
580	Db		639
581	Db		640
582	Db		641
583	Db		642
584	Db		643
585	Db		644
586	Db		645
587	Db		646
588	Db		647
589	Db		648
590	Db		649
591	Db		650
592	Db		651
593	Db		652
594	Db		653
595	Db		654
596	Db		655

Mon Aug 16 10:21:24 1999

transcriptase"

40: 40 cm from centromere

[illegible][illegible]

Db 1561 GCAGCAGCCGGGAGGAGCCGTGTCCTCCGCTGCAGAGCACCGTCTGAGGAGAGGATCC 1620  
Qy 1630 TGGCTACGTTCTGTTCTGGCTGATGGACACATAGCTGCTACAGCTGCTTAGGTCATTCT 1689  
Db 1621 TGGCTACGTTCTGTTCTGGCTGATGGACACATAGCTGCTTAGGTCATTCT 1690  
Qy 1690 TTTACATCACAGAGAGACATTTCCAGAGAGACAGGCTCTTCTTACCGTAAGAGTGTG 1749  
Db 1681 TTTACATCACAGAGAGACATTTCCAGAGAGACAGGCTCTTCTTACCGTAAGAGTGTG 1740  
Qy 1750 GGAGCAAGCTCGAGAGCATTTGAGTTCAGGCAACACCTTGAGAGAGTGGCTAGGGAGC 1809  
Db 1741 GGAGCAAGCTCGAGAGCATTTGAGTTCAGGCAACACCTTGAGAGAGTGGCTAGGGAGC 1800  
Qy 1810 TGTCAAGAGAGAGTCAAGGATCACAGAGACACCTGGCTAGGCATGCCATCTGCAGAC 1869  
Db 1801 TGTCAAGAGAGAGTCAAGGATCACAGAGACACCTGGCTAGGCATGCCATCTGCAGAC 1860  
Qy 1870 TGGCTTCATCCCAAGCCCAAGCCCTGGGCCCATTTGCTGCTGCTGCTGCTGCTGCTG 1929  
Db 1861 TGGCTTCATCCCAAGCCCAAGCCCTGGGCCCATTTGCTGCTGCTGCTGCTGCTGCTG 1920  
Qy 1930 GTACCAGAGCTTTGGGAG 1989  
Db 1921 GTACCAGAGCTTTGGGAG 1980  
Qy 1990 TCTTCAGCATGCTCAACTATGAGCGGCAAAACATCTTCACTTATGGGGTCTTCTGTAT 2049  
Db 1981 TCTTCAGCATGCTCAACTATGAGCGGCAAAACATCTTCACTTATGGGGTCTTCTGTAT 2040  
Qy 2050 TGGGTATGAATGACATCTACAGGACCTGGCGGCCCTTTGCTGCTGCTGCTGCTGCTG 2109  
Db 2041 TGGGTATGAATGACATCTACAGGACCTGGCGGCCCTTTGCTGCTGCTGCTGCTGCTG 2100  
Qy 2110 ACCAGACACCCAGAGTGTACTTTGTTAAGGCAGATGTGACCGGGGCTATGATGCCATCC 2169  
Db 2101 ACCAGACACCCAGAGTGTACTTTGTTAAGGCAGATGTGACCGGGGCTATGATGCCATCC 2160  
Qy 2170 CCCAGGGTAACTGCTGAGGAGTGTGTTGCCAATATGATCAGGACCTCGGAGACAGTACT 2229  
Db 2161 CCCAGGGTAACTGCTGAGGAGTGTGTTGCCAATATGATCAGGACCTCGGAGACAGTACT 2220  
Qy 2230 GTATCCGCCAGATGACAGTGTGCTGCGAGAGATAGCCAGGCCCAAGTCCCAAGTCTTTA 2289  
Db 2221 GTATCCGCCAGATGACAGTGTGCTGCGAGAGATAGCCAGGCCCAAGTCCCAAGTCTTTA 2280  
Qy 2290 GGAGACAGTCAACACCTCTCTGACCTCCAGCCATACATGGCCAGTTCCTTAAGCATC 2349  
Db 2281 GGAGACAGTCAACACCTCTCTGACCTCCAGCCATACATGGCCAGTTCCTTAAGCATC 2340  
Qy 2350 TGCAGGATTCAGATGCCAGTGTGCTGAGGAACTCCGTTGTATCAGCAGAGCATCTCTA 2409  
Db 2341 TGCAGGATTCAGATGCCAGTGTGCTGAGGAACTCCGTTGTATCAGCAGAGCATCTCTA 2400  
Qy 2410 TGAATCAGAGCAGCAGCCTGTTGACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 2469  
Db 2401 TGAATCAGAGCAGCAGCCTGTTGACTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 2460  
Qy 2470 TAAAGATTGGTACAGTGTCTATAGCAGTGGCAGGATCCCGGAGGATCCCGGAGGCTAT 2529  
Db 2461 TAAAGATTGGTACAGTGTCTATAGCAGTGGCAGGATCCCGGAGGATCCCGGAGGCTAT 2520  
Qy 2530 CCACCTGCTCTGCAGTCTGTGTTTCGAGACATGGAGAAACAGCTGTTGCTGAGGTGC 2589  
Db 2521 CCACCTGCTCTGCAGTCTGTGTTTCGAGACATGGAGAAACAGCTGTTGCTGAGGTGC 2580  
Qy 2590 AGCGGATGGGTTGCTTTTACGTTTGTGATGACTTCTGTTGGTACGCCCTCACTTGG 2649  
Db 2581 AGCGGATGGGTTGCTTTTACGTTTGTGATGACTTCTGTTGGTACGCCCTCACTTGG 2640  
Qy 2650 ACCAAGCAAAACCTTCTCAGCACCCTGGTCCATGCGCTTCTGATGATGGTGCATGA 2709  
Db 2641 ACCAAGCAAAACCTTCTCAGCACCCTGGTCCATGCGCTTCTGATGATGGTGCATGA 2700

Qy 2710 TAACTTGCAGAGACAGTGTGTAACCTTCCCTGTGGAGCTGGTACCCTGGTGGTGAC 2769  
Db 2701 TAACTTGCAGAGACAGTGTGTAACCTTCCCTGTGGAGCTGGTACCCTGGTGGTGAC 2760  
Qy 2770 CTCCATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2829  
Db 2761 CTCCATACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2820  
Qy 2830 AGACTTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2889  
Db 2821 AGACTTGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2880  
Qy 2890 TCACCTTCCAGAGTGTCTTCAAGCTGGGAAGACCATGCGGAACAAGCTCTGCTGCTG 2949  
Db 2881 TCACCTTCCAGAGTGTCTTCAAGCTGGGAAGACCATGCGGAACAAGCTCTGCTGCTG 2940  
Qy 2950 TCGGTTGAAAGTGTACAGCTGCTTATTTCTAGACTTGCAGGCTACAGGTTCCATGCTG 3009  
Db 2941 TCGGTTGAAAGTGTACAGCTGCTTATTTCTAGACTTGCAGGCTACAGGTTCCATGCTG 3000  
Qy 3010 GCATCAATATATACAAGATCTTCTGCTTCAAGCTACAGGTTCCATGCTGCTGCTG 3069  
Db 3001 GCATCAATATATACAAGATCTTCTGCTTCAAGCTACAGGTTCCATGCTGCTGCTG 3060  
Qy 3070 AGCTTCCCTTTGACCGGCTGTTAGGAAGAACCTCAGATCTTCTTCTGGGATCATCTCA 3129  
Db 3061 AGCTTCCCTTTGACCGGCTGTTAGGAAGAACCTCAGATCTTCTTCTGGGATCATCTCA 3120  
Qy 3130 GCCAAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCA 3249  
Db 3121 GCCAAGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCA 3240  
Qy 3190 CCTGCTGCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCA 3309  
Db 3181 CCTGCTGCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCA 3300  
Qy 3250 AGCTGCTGCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCA 3369  
Db 3241 AGCTGCTGCTCTTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTCA 3360  
Qy 3310 AACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3429  
Db 3301 AACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3420  
Qy 3370 CAGCCCTTAAGCAGACACTTTTCCAGACCATTTTCCAGACTAACCTGCTGCTTCCGCTAGATG 3429  
Db 3361 CAGCCCTTAAGCAGACACTTTTCCAGACCATTTTCCAGACTAACCTGCTGCTTCCGCTAGATG 3420  
Qy 3430 AACATG 3435  
Db 3421 AACATG 3426

## RESULT 2

AF073311 3369 bp mRNA ROD 09-SEP-1998  
LOCUS Mus musculus telomerase catalytic subunit mRNA, complete cds.  
DEFINITION AF073311  
ACCESSION 93551846  
NID AF073311.1 GI:3551846  
VERSION  
KEYWORDS house mouse.  
SOURCE Mus musculus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3369)  
AUTHORS Martin-Rivera, L., Herrera, E., Albar, J.P. and Blasco, M.A.  
TITLE Expression of mouse telomerase catalytic subunit in embryos and adult tissues  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (18), 10471-10476 (1998)  
MEDLINE 98393668  
REFERENCE 2 (bases 1 to 3369)

AUTHORS Martin-Rivera, L., Herrera, E. and Blasco, M.A.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUN-1998) Immunology and Oncology, National Centre of  
Biotechnology, Cantoblanco, Madrid 28049, Spain

FEATURES Location/Qualifiers  
source 1. .3369  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
1. .3369

CDS  
/codon\_start=1  
/product="telomerase catalytic subunit"  
/protein\_id="AAC34821.1"  
/db\_xref="PID:3551847"  
/db\_xref="GI:3551847"  
translation="MTRAPCPAVRSLRSRYREVPWVPLATFVRRLGPRRLVOPGP  
KIVRLVLCVLMHWSQPPADLSFHQVSKLDELVARVVRVORCERNERNVLAAGE  
LINEARGPMAFTSVRSYLPNTVETLRVSGAWMLLSRGDDLLVYLALHACALYL  
LVPPSAYOCVGPLYICATTDIWPVSASACYPVPRVEGPHROVLPTPSGKSWPSP  
AKPALPRGKRLHLSLSTSVPSAKARCYVPVVEGPHROVLPTPSGKSWPSP  
ARSPVPTAEKLDSSKGVSDLSGSCVCKHKPSSTLSLPPRONALORLPFTETRH  
FLYSGDGOERLNPFLSNLQPNLITGARRLVEILFLGSRPTSGPLORHLRRLRY  
QMRPLQQLVNHAEQYVRLLRSHCFRTANQVTDALNTSPHMLDLRLHSPWQ  
WGLFRLACILKVVASLWGTNRHREFFKLFISLGYKLSLQELMWMKVEDCH  
WLRSSPKDRVPAEHLRLERILATFELWMDTYVQLRFFYIETEFQKRLFFY  
RKSVMKLSIGVROHLEHRLVRELSEVHRHQTWILAMPICRLRFPKNGRLPIV  
FVLRALDGTQPMYFVKADVTCAYDAIPOGLVENVANNIRHSESTYCIROYAVWRR  
NSYSXGTRALGRKQAOHETORLKTFLSMLNVERTKPHLMGSSVLGMDIYRTWRA  
DSQGVHFRVOTLSDLOPNQFGLAHLODSASALRNSVLEQSIIMNESSL  
FDFLHFLRHSVYKIDRCYTCOGIPQSSSLTLCISGICFCFDMENKLEAFVORGLL  
LRFVDDPLVLPHTDQAKTFLSLVHVEYQGMINLQTVNVPFVPEVGTGGAAPYQ  
LPACILFPWGLLDQTLQTEVFDYQYATSIKTLTQSFVKAGKMRNKLSSYLR  
LKCHGLFDLVQNSLQVTCINIKYIFLQAYRPHACVIOIPFDORVKNLTFFLGIIIS  
SQASCCVAILKKNPMTLAKSGFPEAAHWCYQAFLLKLAHSVYVKLLGLPLRT  
AQKLCRLPEATWILKAADPAISDFQITLD"

BASE COUNT 737 a 959 c 891 g 782 t  
ORIGIN

Query Match 96.3%; Score 3367.4; DB 12; Length 3369;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 3368; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 39 ATGACCCGCGCTCTCGTTCGCCGCGGTCGCTCTCTGCTGCGCAGCCGATACCGGGAG 98  
Db 1 ATGACCCGCGCTCTCGTTCGCCGCGGTCGCTCTCTGCTGCGCAGCCGATACCGGGAG 60  
Qy 99 GTGTGCGCGCTGCAACCTTTGTGCGGCGCTTGTGCGGCGCGGAGCGCGCTTTGTGCAA 158  
Db 61 GTGTGCGCGCTGCAACCTTTGTGCGGCGCTTGTGCGGCGCGGAGCGCGCTTTGTGCAA 120  
Qy 159 CCGGGGACCCGAGATCTACCGCATTTGTGTTGCCCAATGCTAGTGTGATGCACCTGG 218  
Db 121 CCGGGGACCCGAGATCTACCGCATTTGTGTTGCCCAATGCTAGTGTGATGCACCTGG 180  
Qy 219 GGCTACAGCTCCAGCTGCGGACCTTTCTTCCACAGGTGTCTCCCTGAAAGAGCTG 278  
Db 181 GGCTACAGCTCCAGCTGCGGACCTTTCTTCCACAGGTGTCTCCCTGAAAGAGCTG 240  
Qy 279 GTGGCAGGGTGTGCGAGACTCTGCGAGCGCAACGAGAGAAACGCTGCTGCTTTGGC 338  
Db 241 GTGGCAGGGTGTGCGAGACTCTGCGAGCGCAACGAGAGAAACGCTGCTGCTTTGGC 300  
Qy 339 TTTGAGTGTCTTAACGAGGCGAGCGGCGCTCCCTTCCAGCTTCACTAGTACGCTGGCT 398  
Db 301 TTTGAGTGTCTTAACGAGGCGAGCGGCGCTCCCTTCCAGCTTCACTAGTACGCTGGCT 360  
Qy 399 AGCTACTTGCCCAACACTGTTATAGACACCTCGCTGTGTCAGTGGTGCATGGATGCTACTG 458  
Db 361 AGCTACTTGCCCAACACTGTTATAGACACCTCGCTGTGTCAGTGGTGCATGGATGCTACTG 420  
Qy 459 TTGAGCGGAGTGGGCGAGGACCTGCTGCTACTGCTGCGACACTGTGCTCTTTATCTT 518  
Db 421 TTGAGCGGAGTGGGCGAGGACCTGCTGCTACTGCTGCGACACTGTGCTCTTTATCTT 480

Qy 519 CTGGTCCCGCCAGCTGTGCTTACCAGGTGTGGTGTCCCGCTTACCAAAATTTGTGCC 578  
Db 481 CTGGTCCCGCCAGCTGTGCTTACCAGGTGTGGTGTCCCGCTTACCAAAATTTGTGCC 540  
Qy 579 ACCACGGATATCTGGCCCTCTGTGTCCCGTGTAGTACAGGCCCCACCGCCGCTGGGAGG 638  
Db 541 ACCACGGATATCTGGCCCTCTGTGTCCCGTGTAGTACAGGCCCCACCGCCGCTGGGAGG 600  
Qy 639 AATTTCCTAACTTAGCTTCTTACCAAGAGTCAAGAGCAGTAGTGCAGGAAGACCG 698  
Db 601 AATTTCCTAACTTAGCTTCTTACCAAGAGTCAAGAGCAGTAGTGCAGGAAGACCG 660  
Qy 699 AAACCCCTGGCCCTTGCATCTCGAGGTACAAAGAGGATCTGAGTCTCACCAGTACAAGT 758  
Db 661 AAACCCCTGGCCCTTGCATCTCGAGGTACAAAGAGGATCTGAGTCTCACCAGTACAAGT 720  
Qy 759 GTGGCTTACAGTAAAGAGGCGCAGATGCTATCTCTCCGAGAGTGGAGGAGGAGGACCCAC 818  
Db 721 GTGGCTTACAGTAAAGAGGCGCAGATGCTATCTCTCTTAAAGAAAGGTGTCTGACCTGAGTCTC 900  
Qy 819 AGGAGGTGTGTACCAACCCCATCAGGCAAAATCATGGGTGCCAAGTCTGCTCGGTCCCC 878  
Db 781 AGGAGGTGTGTACCAACCCCATCAGGCAAAATCATGGGTGCCAAGTCTGCTCGGTCCCC 840  
Qy 879 GAGTGCTCTACGAGAAAGATTTCTCTTAAAGAAAGGTGTCTGACCTGAGTCTC 938  
Db 841 GAGTGCTCTACGAGAAAGATTTCTCTTAAAGAAAGGTGTCTGACCTGAGTCTC 900  
Qy 939 TCTGGTGGTGTGTGTGTAAACAAAGCCAGCTTCCACATCTCTGTGTGCACACCCCGC 998  
Db 901 TCTGGTGGTGTGTGTGTAAACAAAGCCAGCTTCCACATCTCTGTGTGCACACCCCGC 960  
Qy 999 CAAATGCTTTAGCTCAGGCCATTTATTGAGACCAAGACATTTCTTACTCCAGGGGA 1058  
Db 961 CAAATGCTTTAGCTCAGGCCATTTATTGAGACCAAGACATTTCTTACTCCAGGGGA 1020  
Qy 1059 GATGCCAAGAGCTTAAACCCCTCATTTCTTACTCAGCAACCTCCAGCTTAAGTCTGACT 1118  
Db 1021 GATGCCAAGAGCTTAAACCCCTCATTTCTTACTCAGCAACCTCCAGCTTAAGTCTGACT 1080  
Qy 1119 GGGGCGAGGAGACTGGTGGAGATCATCTTCTGGGCTCAAGGCTTGGAGATCAGGACCA 1178  
Db 1081 GGGGCGAGGAGACTGGTGGAGATCATCTTCTGGGCTCAAGGCTTGGAGATCAGGACCA 1140  
Qy 1179 CTCTGAGCAGACACCGCTCTATCGGCTGATCTGCTGAGATCGGCGCCCTGTTCACACAG 1238  
Db 1141 CTCTGAGCAGACACCGCTCTATCGGCTGATCTGCTGAGATCGGCGCCCTGTTCACACAG 1200  
Qy 1239 CTGCTGGTGAACCATGCGAGGTGCCAATATGTCAGACTCCTCAGCTCAGTTCAGGTTT 1298  
Db 1201 CTGCTGGTGAACCATGCGAGGTGCCAATATGTCAGACTCCTCAGGTCAGTTCAGGTTT 1260  
Qy 1299 CGAACAGCAACCAACAGGTGACAGATGCTTGAACACAGCCACCGCCACCTCATGAT 1358  
Db 1261 CGAACAGCAACCAACAGGTGACAGATGCTTGAACACAGCCACCGCCACCTCATGAT 1320  
Qy 1359 TTGCTCGGCTGCGACAGCAGTCCCTGGCAGGTATATGTTTCTTCTGGGCTGTCTCTGC 1418  
Db 1321 TTGCTCGGCTGCGACAGCAGTCCCTGGCAGGTATATGTTTCTTCTGGGCTGTCTCTGC 1380  
Qy 1419 AAGTGTGTCTGCTAGTCTCTGGGTACCGACCAATAGAGCGGCTTCTTTAAGAAC 1478  
Db 1381 AAGTGTGTCTGCTAGTCTCTGGGTACCGACCAATAGAGCGGCTTCTTTAAGAAC 1440  
Qy 1479 TTAAGAGTTCATCTCGTTGGGAAATACGCAAGCTATCTGCTGAGAACTGATGTGG 1538  
Db 1441 TTAAGAGTTCATCTCGTTGGGAAATACGCAAGCTATCTGCTGAGAACTGATGTGG 1500  
Qy 1539 AAGTGAAGTAGAGGATTCGCTGCTCCGAGGAGCCCGGGAGGACCGGTGTCCCC 1598  
Db 1501 AAGTGAAGTAGAGGATTCGCTGCTCCGAGGAGCCCGGGAGGACCGGTGTCCCC 1560





[illegible]



Mon Aug 16 10:21:24 1999

Best Local Similarity 68.9%; Pred. No. 0; Mismatches 967; Indels 108; Gaps 9; Matches 2379; Conservative

Db 4452 GCTGAGTCGGAGCTCCCGGAGCAGCTGACTGCTGAGGCGCGACCAACCGGC 4511  
 QY 3374 CTAAGCAGACACTTTCAGACATTTGGAGCAACCTGCTCTCTCC 3421  
 Db 4512 ACTGCCCTCAGACTTCAAGACCATCTGGACTACCCATACAGATGCC 4559

RESULT 4  
 AF015950 4015 bp mRNA PRI 16-AUG-1997  
 LOCUS Homo sapiens telomerase reverse transcriptase (hTRT) mRNA, complete  
 DEFINITION cds.  
 ACCESSION AF015950  
 NID 9230016  
 VERSION AF015950.1 GI:2330016  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 4015)  
 Nakamura, T.M., Morin, G.B., Chapman, K.B., Weinrich, S.L.,  
 Andrews, W.H., Lingner, J., Harley, C.B. and Cech, T.R.  
 Telomerase catalytic subunit homologs from fission yeast and human  
 Science 277 (5328), 955-959 (1997)  
 97400623  
 2 (bases 1 to 4015)  
 Morin, G.B.  
 Direct Submission  
 Submitted (24-JUL-1997) Genon Corporation, 230 Constitution Drive,  
 Menlo Park, CA 94025, USA  
 Location/Qualifiers  
 1..4015  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue\_type="kidney"  
 /map="5p"  
 /dev\_stage="embryo"  
 /chromosome="5"  
 1..4015  
 /gene="hTRT"  
 56..3454  
 /gene="hTRT"  
 /codon\_start=1  
 /product="telomerase reverse transcriptase"  
 /protein\_id="AAC51672.1"  
 /db\_xref="PID:9230017"  
 /translation="MPRAPRCRAVRLLSHRYEVLPLATFVRRLGPGWRLVQGGDP  
 AAFRLVAQCLVCPWDARPPAPPSFROVSLKELVALVRLQRCERGAKNVLAFFGA  
 LLDGARGGPEAFTSVSYLPNTVDALRGSGANGLLLRVGGDDVLVHLLARLAFV  
 LVASCAVQCGPPLYLQGAATQARPPEPHASGFRKLGCERANWHSREAGVPLGLPA  
 PGARRGSASRLPLPKRPRGAAPPERTPVGGSWAHPGRTKPSDRTGFCVVSFA  
 PBAEATSLGALSSTHSHPSVGRQHHAGPPSTSRPPRPMTCPPPVIAETKHFLYS  
 SGDEQLRPSALLSLRPSLTGARLVETIFLGSRPMPGTPRPLRPQRYQWRPL  
 FLELLNHAQCPYGVLLKTHCPLEAATPAAGVAREKPGQSVAAPEEDIDPRLVQ  
 LLRQSSHPQYGVFRACLRLVPLGLWSHNERFLRLTKFIISLGRHAKLSLOEL  
 TWKMSVRCAWLKRSVGVCPVAAEHLREELAKFLHMLSVVYVELLSFFVYET  
 TKMRLFFYKRSVMSKLSIGIFQHLRLVQLRELSAEVROHREARPLLSRLFI  
 PPDGLFVNNVYVVGARTFRREKARELTSRKALFVNLVETIASILKPNQYTC  
 LDDIHRAPFTVLRVADQPPPELYFKVDVTGAYDTPDRLTEVIAISILKPNQYTC  
 LVRIAYVKAAGHVRKAFKSHVSTLDLPYMRQFVAHLOETSPRLDAVVIDQSSSL  
 NEASGLDFVLRFPCHAVRTKGSVYVOCQIPQSGISLTSLCSLGDWENKLEAF  
 IRDGLLLRLHDFLLVTPHLLTARLTLVLRVGVGVNLRKTVNFRVFNVEDAL  
 GGTAFVLRKCHSLFDLQVNSLQTVCTNLYKLLQVAFRHACVLOQFPFHQYWKNP  
 KFLGVLRKCHSLFDLQVNSLQTVCTNLYKLLQVAFRHACVLOQFPFHQYWKNP  
 YFVLGSLRTAQTQLSRKLPGLTTLTAAANPALPSDFKTLID"  
 663 a 1363 c 1275 g 714 t

BASE COUNT 663 a 1363 c 1275 g 714 t  
 ORIGIN

Query Match 45.8%; Score 1600.8; DB 11; Length 4015;

23 CCGGGCTTTGAGCACAATGACCCCGCTCTCTGTTGCCGCGCTCTCTCTGTCG 82  
 40 CCGGCGACCCCGGATGCGCGCGCTCCCGCTGCGAGCCGCTCTCTCTGTCG 99  
 83 CAGCCATACCGGAGGTGTCGCTGTCGCAACCTTTGTGGGCGCTTGGGCGCCGAGG 142  
 100 CAGCCATACCGGAGGTGTCGCTGTCGCAACCTTTGTGGGCGCTTGGGCGCCGAGG 159  
 143 CAGCGCTTTGTCAACCCCGGAGCCGCAATCTACGCACTTTGGTTGCCAATGCT 202  
 160 CTGCGCTGTCGAGCGCGGAGCCCGCGCTTCCGCGGCTGTTGGCCAGTGCCT 219  
 203 AGTGTGATGCTGCGGCTCACAGCTCTCACTGCGGACCTTTCTTCCACAGGTGTC 262  
 220 GGTGTGCTGCTGCGGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 279  
 263 ATCCCTGAAGAGTGTGGCGGAGGTGTGAGAGCTCTGCGAGCGCAACAGAGAAA 322  
 280 CTGCTGAAGAGTGTGGCGGAGGTGTGAGAGCTGTGAGAGCTGTGAGAGCGGCGG 339  
 323 CGTGTGCTTTTGGCTTTGAGCTGCTTAACAGGCGCAGAGCGGCGGCTTCCATGCTT 382  
 340 CGTGTGCTTTTGGCTTTGAGCTGCTTAACAGGCGCAGAGCGGCGGCTTCCATGCTT 399  
 383 CACTAGTACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 442  
 400 CACCACAGCGTGGCGAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 459  
 443 TGCATGATGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 502  
 460 GCGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 519  
 503 CTGTGCTTTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 562  
 520 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 579  
 563 GTACCAATTTGTGCCACCGGATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 622  
 580 GTACCACTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 639  
 623 CCGACCTGCGGAGGAAATTTACTAATCTAGTTTCTTACAAACAGATCAAGACGATAG 682  
 640 GCGTCTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 672  
 683 TCGCCAGGAGCAGCAACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 742  
 673 CGTACGAGGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 732  
 743 TCTCAGCTACAAGTGTGCTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 802  
 733 CAGTGCAGCGGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 792  
 803 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 862  
 793 GGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 852  
 863 TCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 913  
 853 TGACCGTGGTTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 912  
 914 AGGAAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 970  
 913 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 972  
 971 CTCACAT---CTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1027  
 973 CCCCCCATCCATCGCGCGGAGCCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1032  
 1028 TGAGACCAAGATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1087

Db 1033 CGAGACCAAGCACTTCTCTACTCTCAGCGA---CAAGGAGCAGCTGCGGCCCTCCTT 1089  
Qy 1088 CCTACTCAGCAACCTCCAGCCTAACTTGTGAGTGGGCCAGGAGACTGTGTGGAGATCATCTT 1147  
Db 1090 CCTACTCAGCTCTCTGAGGCCAGCCTGACTGCGCTCGAGGCTCTGTGGAGACCATCTT 1149  
Qy 1148 TCTGGGCTCAAGGCTTAGGACATCAGGACCACTCTGAGGACACACCGCTATCGGGTGG 1207  
Db 1150 TCTGGGCTTCCAGGCTTGATGCGCAGGACTCCCGGAGGTTGCGCGCTGCCAGCG 1209  
Qy 1208 ATACTGGCAGATGCGGCCCTTGTCCAAACAGTCTGTGTGAACATGCGAGTGGCAATA 1267  
Db 1210 CTACTGGCAATGCGGCCCTTGTCTGGAGCTCTTGGGAACCGAGCGAGTGGCCCTA 1269  
Qy 1268 TGTCAAGCTCTCAGGTCACATTCAGGTTTCCGAACAGCAACCAAGGTGACAGATGC 1327  
Db 1270 CGGGGTGCTCTCAAGACGCACTGCCGCTCGAGCTGGGTGACCCAGCAGCGGGT 1329  
Qy 1328 CT-----TGACACCAAGCCC 1342  
Db 1330 CTGTGCGCGGAGAACCCAGGGCTCTGTGCGGCCCGCCGAGGAGGAGACAGACCC 1389  
Qy 1343 ACCGACCTCATGATTTGCTCGCTGCACAGCAGTCCCTGGCGAGTATATGTTTCT 1402  
Db 1390 CCGTGGCTGTGACGTGCTCGCGCAGCAGCAGCCCTGGCAGGTGTACGGCTTCT 1449  
Qy 1403 TCGGGCTGTCTCTCAAGGTGGTGTCTGCTAGTCTCTGGGTACAGGACACATGAGCG 1462  
Db 1450 GCGGGCTGCTCGCGGCTGGTCCCGCCAGGCTCTGGGGCTCCAGGACACCAAGCG 1509  
Qy 1463 CGCTTCTTAAAGACTTAAAGTTCATCTCTGTGGGGAATACGGCAAGTATCACT 1522  
Db 1510 CGCTTCTCAAGAACCAAGAGTTCATCTCTCGGGAGCATGCCAAGCTCTCGCT 1569  
Qy 1523 GCAGGAAGTGTGGAAGATGAAGTAGAGGATTCCTGCTCGGAGAGTCTGAGGAGTCTT 1642  
Db 1570 GCAGGAGTGTGAGGAGTGTGAGGAGTGTGCGGAGTGTGCTGCGCAGGAGCCAGG 1629  
Qy 1583 GAAGGACCGTCTCCCGCTCGAGGACCGCTGTGAGGAGAGGATCCCTGGCTACGTTCT 1642  
Db 1630 GGTGGCTGTCTCGGCCCGCAGAGACCGCTGTGAGGAGTCTGAGGAGTCTTGGCAAGTCT 1689  
Qy 1643 GTTGTGCTGATGGACACATACCTGTGTACAGTGTCTAGGTCATTTTACATCAGCA 1702  
Db 1690 GCACTGTGCTGATGAGTGTGTACGTGTGAGTGTCTGAGTGTCTGAGTGTCTGAGG 1749  
Qy 1703 GAGCACATTCCAGAAACAGGCTCTTCTTACCTGAGGAGTGTGAGTGTCTGAGTGTCT 1762  
Db 1750 GACCACTTCAAGAAACAGGCTCTTCTTACCGGAGAGTGTCTGAGGAGTGTCTGAG 1809  
Qy 1763 GACATTGGAGTCAGGCAACACTTGACAGAGTGTGCGGTACGGAGTGTGTCAAGAGGA 1822  
Db 1810 AAGCATTTGAATCAGACGACTTGAAGAGGTGCAGTGTGCGGAGTGTGTGGAAGCAGA 1869  
Qy 1823 GGTGAGGATACCAAGGACCTGGCTAGCCATGCCATCTGAGACTGGCTTATGAGTGTG 1882  
Db 1870 GGTGAGGACATCGGAAGCAGGCGCCCTGCTGACCTCGAGACTCGGCTTATGCC 1929  
Qy 1883 CAAGCCCAACCGCTGCGGCCCTTGTGAACATGATTTAGCATGGGTACCAAGCTTT 1942  
Db 1930 CAAGCCTGACGGCTGCGCGGATTTGAACATGAGTGTGAGTGTGAGTGTGAGTGTG 1989  
Qy 1943 GGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2002  
Db 1990 CCGCAGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2049  
Qy 2003 CAACATGAGCGGACAAAATCTCCTCACCCTTATGGGTCTTGTACTGGGTATGAATGA 2062  
Db 2050 CAATAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2109  
Qy 2063 CATCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2122  
Db 2110 TATCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2169

Qy 2123 GATGTACTTTTAAAGCAGATGTGACCGGGCCCTATGATGCCATCCCCAGGTAAGCT 2182  
Db 2170 GCTGTACTTTGTCAAGTGATGTGACGGCGCGTACGACACCATCCCCAGGACAGGCT 2229  
Qy 2183 GGTGGAGGTTGTGCAATATATGATCAGGCACTCGGAGAGCAGCTACTGTATCCGCCAGTA 2242  
Db 2230 CAGGAGGTCATCGGCAGCATCATCA---AACCCAGAACACAGTACTGCGTGGTGGTA 2286  
Qy 2243 TGCAGTGGTCCGAGAGATAGCCAGGCCCAAGTCCCAAGTCTCTTTAGGAGACAGTGCAC 2302  
Db 2287 TGCCTGTGTCAGAGAGCGGCCCATGGCAGCTCGCAAGCCCTTCAAGAGCCAGCTCTC 2346  
Qy 2303 CACCCTCTCTGACCTCCAGCCATACATGGCCAGTTCCTTTAAGCATCTCGAGATTCAGA 2362  
Db 2347 TACCTTGACAGACCTCCAGCGCTACATGCGACAGTTCTGGTCTCAGCTGAG-----GA 2400  
Qy 2363 TGCCAGTGCATGAGGAATCCGTTGTTCATCGAGCAGCATCTCTATGAATGAGAGCAG 2422  
Db 2401 GACCAGCCGCTGAGGATGCCCTGTCATCGAGCAGAGCTCCTCCTGAATGAGGCGAG 2460  
Qy 2423 CAGCAGCTCTTTGACTTCTCTGCACTTCTGCTGCTCACAGTGTCTGTAAGATTTGTTGA 2482  
Db 2461 CAGTGGCTCTTCCGAGCTTCTTACGCTTCATGTGCCACACCGCTGCGGCGAGCGGCT 2520  
Qy 2483 CAGGTGCTATPACGAGTGCAGGATCCCGCAGGGCTCCAGGCTATCCACCTGCTCTG 2542  
Db 2521 CAAGTCTCTAGTCCAGTGCAGGGATCCCGAGGGTCTCCTCTCCACGCTGCTCTG 2580  
Qy 2543 CAGTCTGTCTTCGAGACATGGAGAACAGCTGTTTGTCTGAGTGTGAGGAGGATGGGTT 2602  
Db 2581 CAGCTGTGCTACGCGGACATGGAGAACAGCTGTTTGTGCGGGATTCGGCGGAGCGGCT 2640  
Qy 2603 GCTTTTACGTTTGTGATGACTTCTTGTGTGACGCTCCTTGGACCAAGCAAAAC 2662  
Db 2641 GCTCTGTGCTTGTGATGATTTCTTTGTTGACACCTCACCCTCAGCCAGCGGAGGCT 2700  
Qy 2663 CTCTCTACGACACCTGTGCTCCATGCGTTCCTGATATGGGTGCATGATAAAGTTCAGAA 2722  
Db 2701 CTCTCTCAGGACCTGTGCGAGGTGCTGCTGATGATGCTGCTGCTGCTGCTGCTGCT 2760  
Qy 2723 GACAGTGTGAACTTCTCTGAGGCTGCTGCTGAGTGTGCTGCTGCTGCTGCTGCTGCT 2782  
Db 2761 GACAGTGTGAACTTCTCTGAGGCTGCTGCTGAGTGTGCTGCTGCTGCTGCTGCTGCT 2820  
Qy 2783 GCTGTCTGCTGCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2842  
Db 2821 GCGGGCCAGGCTATCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2880  
Qy 2843 GTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2902  
Db 2881 GCAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2940  
Qy 2903 TGTCTTAAAGTGTGAGACCATGCGGACACAGCTCCTGCTGCTGCTGCTGCTGCTGCT 2962  
Db 2941 GCGCTTCAAGCTGGGAGACATGCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3000  
Qy 2963 TCAGGCTGCTTCTGAGCTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3022  
Db 3001 TCAGGCTGCTTCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGT 3060  
Qy 3023 CAAGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3082  
Db 3061 CAAGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3120  
Qy 3083 CCAGCGTGTGAGGAGACCTCAGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3142  
Db 3121 TCAGGAGTGTGAGGAGACCTCAGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3180  
Qy 3143 CTGCTATGCTTCTGAGGCTCAAGATCCAGGATGAGATGAGTGTGAGTGTGAGTGTGAG 3193  
Db 3181 CTGCTATGCTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3240







us-09-042-460-1.1.rge

Mon Aug 16 10:21:24 1999

```
/db_xref="PID:g4580665"
/db_xref="GI:4580665"
/translation="MTRAPRCPPAVRSLLRSYREVWPLATFVRRLLPGEGRRLVQPGDP
KIYTLVAOCLVCMHWSQPPADLSHOVSLKELVARVYQRCERNENLVAFGE
LLNEARGPMPMAFTSVRSILPNVETLRYSGAMLLSLRVGDDLLVYLIAICALYL
LVPPCAYQVCGSPYLIQIATDIPWSASVSRTPVGRNFTNRFLOQIKSSRQE
AKPLALPSRGTKRHLSTSTSPVSKAKARCPVPRVEGPHROVLPTPGSKSWPSP
ARSPVPTAKDLSSKGVSDLSLGSVCCRHKPSSTSLSPRONAFOLRPFIERRH
FLYSGDGERLNPFLSNLQNLGARRLVEILFSGSRPTSGPLCRTHRLSRVW
OMRFLFOQLLVNHAECQYVRLRSHCRFTANQVQVTDALNTSPHMLDLRLHSSPWQ
VYGLRACLCCKVYSASLWGRNRRFRFNKLFISLKYKGLSLQELMWMKVEDCH
WLRSSP"
BASE COUNT      1944 a 1857 c 1815 g 1882 t
ORIGIN

Query Match      41.8%; Score 1462.6; DB 12; Length 7498;
Best Local Similarity 93.2%; Pred. No. 0;
Matches 1583; Conservative 0; Mismatches 14; Indels 102; Gaps 1;

QY 10 GTGGAGGCCCATCCGGCCTTGAGCACAATGACCCGCTCTCTCTCTGTTGCCCGCGGTGC 69
DB 4458 GTGGAGGCCCATCCGGCCTTGAGCACAATGACCCGCTCTCTCTCTGTTGCCCGCGGTGC 4517
QY 70 GCTCTCTGCTGCGCAGCGGATACCGGGAGGTGTGGCGCTGCGCAACCTTTGTGGGGGCC 129
DB 4518 GCTCTCTCTGCTGCGCAGCGGATACCGGGAGGTGTGGCGCTGCGCAACCTTTGTGGGGGCC 4577
QY 130 TGGGGCCCGAGGCGAGCGGCTTGTGCAACCGGGGACCGCAAGATCTACCGCACTTTGG 189
DB 4578 TGGGGCCCGAGGCGAGCGGCTTGTGCAACCGGGGACCGCAAGATCTACCGCACTTTGG 4637
QY 190 TTGCCCAATGCTAGTGTGCATGCACTGGGGCTCAGAGCCTCCACCTGCGGACCTTTCT 249
DB 4638 TTGCCCAATGCTAGTGTGCATGCACTGGGGCTCAGAGCCTCCACCTGCGGACCTTTCT 4697
QY 250 TCC----- 253
DB 4698 TCCACCAAGTGGGCTCCAGGCGGGATCCCATGGGTGAGGGCGGAAACCGGGAGGAC 4757
QY 253 -----ACCAAGTGTGTCATCCC 267
DB 4758 GTGGGATAGTGGCTAGCTCATGTCTAAGACCTCTTCTCTTACCAAGTGTGTCATCCC 4817
QY 268 TGAAGAGCTGTGGCCAGGCTTGTGCAGAGACTGTGGCGGCAACGAGAGAACTGTC 327
DB 4818 TGAAGAGCTGTGGCCAGGCTTGTGCAGAGACTGTGGCGGCAACGAGAGAACTGTC 4877
QY 328 TGGCTTTTGGCTTTGAGCTGCTTAAGAGCCAGAGCGGGCTCCCATGGCTTCACTA 387
DB 4878 TGGCTTTTGGCTTTGAGCTGCTTAAGAGCCAGAGCGGGCTCCCATGGCTTCACTA 4937
QY 388 GTAGCTGCTAGCTACTTGTGCCCAACACTGTATTGAGACCTTGTGCTGTCAGTGTGTCAT 447
DB 4938 GTAGCTGCTAGCTACTTGTGCCCAACACTGTATTGAGACCTTGTGCTGTCAGTGTGTCAT 4997
QY 448 GGATGCTACTTGTAGCCGAGTGGCGGACGACCTGCTGGTCTACCTGCTGGGACACTGTG 507
DB 4998 GGATGCTACTTGTAGCCGAGTGGCGGACGACCTGCTGGTCTACCTGCTGGGACACTGTG 5057
QY 508 CTCCTTTATCTTCTGGTGGCCCCCAGCTGTGCTACCAAGTGTGTGGTGTCTCCCTGTACC 567
DB 5058 CTCCTTTATCTTCTGGTGGCCCCCAGCTGTGCTACCAAGTGTGTGGTGTCTCCCTGTACC 5117
QY 568 AAATTTGTGCCACACGAGATATCTGGCCCTCTGTGTCCGCTAGTTACAGGCCACCGGAC 627
DB 5118 AAATTTGTGCCACACGAGATATCTGGCCCTCTGTGTCCGCTAGTTACAGGCCACCGGAC 5177
QY 628 CCGTGGGAGGAATTTCACTAAGTTCTTACACAGATCAAGAGCAGTAGTCTGCC 687
DB 5178 CCGTGGGAGGAATTTCACTAAGTTCTTACACAGATCAAGAGCAGTAGTCTGCC 5237
QY 688 AGGAGCACCGAAACCCCTTGGCCCTTGGCCATCTCGAGTACAGAGGCGCATCTGAGTCTCA 747
```

```

Db 5238 AGGAAGCACCAGAACCCCTGGCCCTTCGCCATCTCGAGGTACAAAGAGGCATCTGAGTCTCA 5297
Qy 748 CCACTACAAGTGTGCTTCAGTAAGAAGGCCAGATGCTATCTCTGTCGAGAGTGAGG 807
Db 5298 CCACTACAAGTGTGCTTCAGTAAGAAGGCCAGATGCTATCTCTGTCGAGAGTGAGG 807
Qy 808 AGGACCCACAGGAGGTGCTTACCAACCCCATCAGGCAATCATGTTGTCGCAAGTCCCTG 867
Db 5358 AGGACCCACAGGAGGTGCTTACCAACCCCATCAGGCAATCATGTTGTCGCAAGTCCCTG 867
Qy 868 CTCGCTCCCGAGGTGCTTACTCAGAGAAAGATTGTTCTTAAAGGAAAGGTGCTG 5417
Db 5418 CTCGCTCCCGAGGTGCTTACTCAGAGAAAGATTGTTCTTAAAGGAAAGGTGCTG 927
Qy 928 ACCTGAGTCTCTCTGGGTGCGTGTGCTTAACACAGAGCCAGCTCCACATCTCTGCTGT 5477
Db 5478 ACCTGAGTCTCTCTGGGTGCGTGTGCTTAACACAGAGCCAGCTCCACATCTCTGCTGT 987
Qy 988 CACCACCCGCCAAATGCTTTCAGTCAAGGCCATTATTTAGAGACCAACATCTCTGCTGT 5537
Db 5538 CACCACCCGCCAAATGCTTTCAGTCAAGGCCATTATTTAGAGACCAACATCTCTGCTGT 1047
Qy 1048 ACTCCAGGGAGATGGCCAGAGCGCTTAACCCCTCATCTCTACTCAGCAACCTCCAGC 5597
Db 5598 ACTCCAGGGAGATGGCCAGAGCGCTTAACCCCTCATCTCTACTCAGCAACCTCCAGC 1107
Qy 1108 CTAACCTGACTGGGCGAGAGACTGGTGAGATCATCTTCTGCGTCAAGGCCCTAGGA 5657
Db 5658 CTAACCTGACTGGGCGAGAGACTGGTGAGATCATCTTCTGCGTCAAGGCCCTAGGA 1167
Qy 1168 CATCAGGACCACTCTCAGGACACACCGTCTATCGCTGATATGTCAGATGCGGCC 5717
Db 5718 CATCAGGACCACTCTCAGGACACACCGTCTATCGCTGATATGTCAGATGCGGCC 1227
Qy 1228 TGTTCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5777
Db 5778 TGTTCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1287
Qy 1288 ATTGAGGTTTCGACAGCAACCAACAGTGCAGATGCTGCTGCTGCTGCTGCTGCTGCTG 5837
Db 5838 ATTGAGGTTTCGACAGCAACCAACAGTGCAGATGCTGCTGCTGCTGCTGCTGCTGCTG 1347
Qy 1348 ACCTCATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5897
Db 5898 ACCTCATGATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1407
Qy 1408 CCTGCTCTGCAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 5957
Db 5958 CCTGCTCTGCAAGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1467
Qy 1468 TCCTTAAAGACTTAAAGAGTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6017
Db 6018 TCCTTAAAGACTTAAAGAGTTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1527
Qy 1528 AACTGATGTGGAAGATGAAGTAGAGGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6077
Db 6078 AACTGATGTGGAAGATGAAGTAGAGGATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1587
Qy 1588 ACCGTGCTCCCGCTGCAGA 1606
Db 6138 ATGGCTGGTCTCCAGCTGA 6156

RESULT 7
AF114847
LOCUS
DEFINITION Homo sapiens telomerase reverse transcriptase (TEXT) gene, partial
cds
ACCESSION AF114847
NID 94324972
VERSION AF114847.1
KEYWORDS GI:4324972

```

```

SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 12213)
AUTHORS Wu,K.J., Grandori,C., Amacker,M., Simon-Vermot,N., Polack,A.,
Lingner,J. and Dalla-Favera,R.
TITLE Direct activation of TERT transcription by c-MYC
MEDLINE Nat. Genet. 21 (2), 220-224 (1999)
REFERENCE 99140777
AUTHORS 2 (bases 1 to 12213)
Ducrest,A.-L., Amacker,M., Reichenbach,P., Nabholz,M. and
Lingner,J.
TITLE Identification of the human telomerase reverse transcriptase
(htERT) promoter
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 12213)
Amacker,M., Simon-Vermot,N. and Lingner,J.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1998) Telomerase Research Unit, ISREC, Chemin de
Boveresses 155, Eplinges 1066, Switzerland
FEATURES
source
1. 12213
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
2. 1158
/rpt_family="L1"
/rpt_type="dispersed"
167..376
/rpt_family="L1"
/rpt_type="dispersed"
384..685
/rpt_family="Alu"
/rpt_type="dispersed"
764..986
/rpt_family="L1"
/rpt_type="dispersed"
3091..3345
/rpt_family="LTR"
/rpt_type="dispersed"
3352..3696
/rpt_family="L1"
/rpt_type="dispersed"
3911..4293
/rpt_family="LTR"
/rpt_type="dispersed"
4487..4787
/rpt_family="Alu"
/rpt_type="dispersed"
5270..5582
/rpt_family="Alu"
/rpt_type="dispersed"
7453..>9129
/gene="TERT"
join(7453..7671,7776..>9129)
/gene="TERT"
/codon_start=1
/product="telomerase reverse transcriptase"
/protein_id="AAD17210.1"
/db_xref="PID:g4324973"
/db_xref="GI:4324973"
/translation="MPAPRCRAVRSLLRSHRYEVLPLATFVRLRPGQWRLVORDDP
AAFRALVAQCLVCFWMDARPEAPSPFQVSKLKLVARVLQRLCERGAKNVLAFGFA
LLDGAGGPFPAFTTSVRSYLPNTVTDALRSGAWGLLRVCGDDVLYHLLARCALFV
LVAPSCATQVCGPLYQLGAATQAPPPHAGSPRRRLGGERANWHSVREAGVPLGLPA
PGARRGGSASRLPLPKRRRGAAPPEPTPVQGSWAHPGTRGSDRGFCVVSFA
PAAEATSLGALSGLTRKSHPSVGRHAGPPTSRPPRPWDTPCPPTVETKHFVLS
SGDKQLRPSFLLSLRPSLTGARRLVETIFLGSREPMPTGTPRRLPQRYWOMRPL
FLELLGHAOCQGVLLKTHCPLRAAVTPAAGVCAAREKPGQSVAAPEEDTPRLVQ
TLRHSVSPQVYGFVRACLRLRVPPLGWSRHNRERFLNKKKISLGHKHLKSLQEL
TWKMSVDCAWLRSP"
7672..7775

```

intron

[illegible]



Mon Aug 16 10:21:24 1999

```

QY 1224 CCCCTGTTCCACAGCTGCTGGTGAACACATGCAGAGTGCACATATATGTCAGATCCTCAGG 1283
Db 14776 CCCCTGTTCTTGGAGCTGCTTGGGAACACAGCGCAGTGCCTTACGGGGTGCCTCTCAAG 14835
QY 1284 TCACATTCAGGTTTCGAACAGCAACCAACAGGTGTACAGATGCTT----- 1331
Db 14836 AGGCACTGCCCGCTCGAGCTGCGGTACCCACAGCGGTGTCTGTGCGCGGAGAAG 14895
QY 1331 -----GAACACGACCCACCGACCTCATGGAT 1358
Db 14896 CCCAGGGCTCTTGGCGSCCCCGAGGAGGACACAGACCCCGCTGCGCTGGTGCAG 14955
QY 1359 TTGCTCGGCTGCACAGAGTCCCTGGCAGGTATATGTTTCTTCGGGCTGTCTCTGC 1418
Db 14956 CTGCTCGGACACAGACGACCCCTGGCAGGTGTACGGCTTCGTGCGGCTGCTCGC 15015
QY 1419 AAGTGTGTCTCTAGTCTCTGGGTACCGACATGAGCGCGCTTCTTTAAGAAC 1478
Db 15016 CGGCTGTGCCCCAGGCTCTGGGCTCCAGGACACAGGACGCGCTTCTCAGGAC 15075
QY 1479 TTAAGAAGTTCATCTGTTGGGAATACGCAAGCTATCACTGCAGGAACATGATGG 1538
Db 15076 ACCAAGAAGTTCATCTCCCTGGGAACATGCCAAGCTCTGCTGCAGGAGTGCAGTGG 15135
QY 1539 AAGATGAAGTAGAGGATTCGCTCCGCTCCGAGCAGCGCGGGAAGACCGTGTCCCC 1598
Db 15136 AAGATGAGCTGCGGAGTTCGCTTGGCTGCGCAGGAGCCAGGTGAGGAGTGTGGCC 15195
QY 1599 GCTGCAGAGCAGCTGTGAGG 1620
Db 15196 GTGAGGAGCCAGGCCCCAGAG 15217

RESULT 9
HSTERT1 HSTERT1 26414 bp DNA PRI 13-MAY-1999
LOCUS Homo sapiens telomerase reverse transcriptase (TERT) gene, exons
DEFINITION 1-6.
ACCESSION AF128893
NID 94808970
VERSION AF128893.1 GI:4808970
KEYWORDS 1 of 2
SEGMENT human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 26414)
AUTHORS Wick.M., Zubov.D. and Hagen.G.
TITLE Genomic organization and promoter characterization of the gene
encoding the human telomerase reverse transcriptase (hTERT)
JOURNAL Gene 232 (1), 97-106 (1999)
MEDLINE 99267414
REFERENCE 2 (bases 1 to 26414)
AUTHORS Wick.M., Zubov.D. and Hagen.G.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-1999) Dept. of Mol. Biol., Bayer AG Leverkusen,
ZF-BTM, Bldg. Q18, Leverkusen D-51368, Germany
FEATURES
source
1. .26414
/organism="Homo sapiens"
/db_xref="taxon:9606"
1. .>11211
/gene="TERT"
promoter
11212. .11492
/gene="TERT"
exon
<number=1
11212. .11273
/gene="TERT"
5' UTR
11597. .12950
/gene="TERT"
exon
<number=2
21762
exon
```

```

/ gene="TERT"
/ number=3
23852. .24032
/ gene="TERT"
/ number=4
24720. .24899
/ gene="TERT"
/ number=5
25394. .25549
/ gene="TERT"
/ number=6
BASE COUNT 5774 a 6788 c 7653 g 6198 t 1 others
ORIGIN
Query Match 13.4%; Score 468.8; DB 42; Length 26414;
Best Local Similarity 59.4%; Pred. No. 8.1e-115; Indels 194; Gaps 7;
Matches 1046; Conservative 0; Mismatches 522;
QY 23 CCGGGCTTGTAGCACAATGACCGGCTCTCTGTTGCCGCGGTGGCTCTCTGTCTGTCTGCG 82
Db 11258 CCGGGCAACCCCGGATGCGCGGCTCCCGCTCCGAGCGTGGCTCTCTGTCTGCG 11317
QY 83 CAGCGGATACCGGAGGTGTGGCGCTGGCAACCTTTGTGCGCGCTGGGCGCCGAGG 142
Db 11318 CAGCACTACCGGAGGTGTGGCGCTGGCAACCTTTGTGCGCGCTGGGCGCCAGG 11377
QY 143 CAGCGGCTTGTGCAACCGGAGCCCGAAGATCTACCGCACCTTTGTTGGCCCAATGCT 202
Db 11378 CTGGCGCTGTGCAACCGGAGCCCGGCGCTTTCCGCGCTGTGTGGCCAGTGCCT 11437
QY 203 AGTGTGATCACTGGGCTCACAGCTCCACCTGCGGACCTTTCTTCCAC----- 255
Db 11438 GTGTGGCTGCTGGGAGCGACGCGCCCGCCCGCCCTCTTCCGCGAGTGG 11497
QY 255 ----- 255
Db 11498 CTTCCCGGGTGGCGTCCGGTGGGTGTAGGGGCGCGGGGGAACACGACATGC 11557
QY 255 -----CAGTGTCACTCCCTGAAGAGCTG 278
Db 11558 GGAGAGCAGCAGCGGACTCAGGGCGCTTCCCGCAGAGTGCTCTGAAGAGCTG 11617
QY 279 GTGGCAGGTTGTGCAGAGACTCTGCAGCGCAACGAGAGAACGTGTGCTTTTGGC 338
Db 11618 GTGGCCGAGTGTGCAGAGCTGTGCAGGGCGCGGAGAACGTGTGCTTGGCTTCGGC 11677
QY 339 TTTGAGCTGCTTAACGAGGCGAGGGCGGCTCCCATGGCTTCACTAGTAGCTGCTGCGT 398
Db 11678 TTCGCGTGTGAGCGGGCGCGCGGGGCGCGCGGCGCTTCACTAGTAGCTGCTG 458
QY 399 AGCTACTTGGCCCAACACTGTTATTGAGACCTGGGTGTGTCAGTGTGATGATGCTACTG 11797
Db 11738 AGCTACTTGGCCCAACACTGTTATTGAGACCTGGGTGTGTCAGTGTGATGATGCTACTG 518
QY 459 TTGAGCGGAGTGGCGACACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTTATCTT 11857
Db 11798 CTGGCGCGGTGGCGACGACGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 578
QY 519 CTGGTGGCGCGCGAGTGTGCTTACAGAGTGTGTGGTGTGCTGCTGCTGCTGCTGCTGCT 11917
Db 11858 CTGGTGGCTCCAGCTGCGCTTACAGAGTGTGCGGGCGCGCGCTGTACAGCTGCGCGCT 578
QY 579 ACCACGATATCTGGCGCTCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 638
Db 11918 GCCACTCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11975
QY 639 AATTTCATACCTTAGGTCTTTACACAGATCAAGAGAGTAGTGTGCGCCAGGACCG 698
Db 11975 -----GAACGCGCTTGAACCATAGCTCAGGAGCGCGCG 12010
QY 699 AAACCCCTGGCTTGCATCTCAGAGTACAAAGAGGATCTGAGTCTCAGCAGTACAAGT 758
```



Db 12011 GTCCCCCTGGCGCTGCAGACCCCGGGTGCAGAGAGCGGGGGCAGTGCCAGCGAAGT 12070  
 QY 759 GTGCTTACCTAAGAGGCGAGATGCTATCTCTCCAGAGAGTGGAGAGGAGCCAC 818  
 Db 12071 CTGCGGTGCCAAGAGGCGAGCGTGCCTGCGCTGAGCGGAGCGAGCGCGCTT 12130  
 QY 819 AGCGAGTGTACCAACCCCATCAGCAATCATGGTGGCCAGTCTGCTGGTCCCCC 878  
 Db 12131 GGCAGGGGTCTGGGCCCCACCCCGGCGAGCGCGTGGACGAGTACCGTGTCTGT 12190  
 QY 879 GAGTGCCTTACT-----CGAGAGAAATTTGTTCTTAAAGAAAGTGTCTGAC 929  
 Db 12191 GTGGTGTACCTGCGAGACCCCGCCGAAGAGCCACTTCTTGGAGGTGCGCTCTGCG 12250  
 QY 930 CTGAGTCTCTC---TGGGTGCGTGTGCTGTAACACAGCCAGCTCCACAT---CTCTG 983  
 Db 12251 AGCGCCACTTCCACCCCATCGTGGCGCCGAGCAGCAGCGAGCCCATCCATCG 12310  
 QY 984 CTGTACCACCCCGCCAAATGCTTTTTCAGCTCAGGCAATTTATTGAGACCGACATTTTC 1043  
 Db 12311 CGGCCACCATGCTGCGGACACGCTTGTCCCGCGTGTACCGCGAGACCAAGCACTTC 12370  
 QY 1044 CTTTACTCAGGAGAGTGGCCAGAGCGCTTAACCCCTCATCTCTACTCAGCAACCTTC 1103  
 Db 12371 CTCTACTCTCAGCGCA---CAAGGACAGCTGCGGCGCTCTCTCTACTCAGCTCTCG 12427  
 QY 1104 CAGCTTACTTGTGCTGGCGGAGAGACTGCTGGAGATCATCTTCTGCGCTCAGGCT 1163  
 Db 12428 AGGCCAGCTGACTGCGCTGCGAGCTGCTGGAGACCATCTTCTGCGTTCAGGGCC 12487  
 QY 1164 AGGACATCAGGACACTCTCAGGACACACCGCTCTATCCGCTGATATCTGCGAGTGG 1223  
 Db 12488 TGGATGCCAGGACTCCCGCAGCTTGCCTGCGCTGCCCGAGCGCTACTGCAATGCG 12547  
 QY 1224 CCCTGTTCACAGCTGCTGTGAACCATGAGTGCATATATGTCAGACTCTCTCAGG 1283  
 Db 12548 CCCTGTTCCTGAGCTGCTGTGGAAACACCGCGAGTGCCTACGCGGTCTCTCTCAAG 12607  
 QY 1284 TCACATTTGAGGTTTCGAACACCAACCAACAGGTGACAGATGCGCT----- 1331  
 Db 12608 ACGCACTGCGCGTGGAGCTGCGTACCCAGCGCGGTGTGTGTCGCGGAGAG 12667  
 QY 1331 -----GAACACACCCACCGACCGACCTATGAT 1358  
 Db 12668 CCGCAGGGCTGTGTGCGCGCCCGGAGGAGGAGGACAGACCGCCGCTGCGCTGGTGCAG 12727  
 QY 1359 TTGCTCCGCTGCAGCAGTCCCTGCGAGATATGTTTTCGCGGCTGTCTCTGC 1418  
 Db 12728 CTGCTCCGCGCAGCAGCGCTTGGCAGGTGTACGGCTTGTGCGGCGCTGCTGCGC 12787  
 QY 1419 AAGTGTGTCTGTAGTCTCTGCGGTACCAAGCACAATGAGCGCGCTCTTTAAGAC 1478  
 Db 12788 CGGCTGTGCCCCCAGGCTCTGGGCTCCAGGCACACCAACAGCGCTTCTCAGGAAC 12847  
 QY 1479 TTAAGAAAGTTTCACTCTGTTGGGAATACGCAAGCTATCATCTGAGGAAGTGTG 1538  
 Db 12848 ACCAAGAGTTTCACTCTCTGCGGAAGCATGCCAAGCTCTGCTGAGGAGTGTG 12907  
 QY 1539 AAGATGAAGTAGAGGATGCCACTGCTGCGCAGCAGCGCGGAGGAGCGTGTCC 1598  
 Db 12908 AAGATGAGCGTGGCGGAGTCTGCTGCGCAGGAGCGGAGGAGGAGTGTGCGC 12967  
 QY 1599 GCTCAGAGCAGCGCTGTGAGG 1620  
 Db 12968 GTCAGGCGCCAGCGCCAG 12989

RESULT 10  
 LOCUS AB016767

DEFINITION Homo sapiens gene for telomerase transcriptase, partial cds.  
 ACCESSION AB016767  
 NID 9423989

LOCUS AB016767 5491 bp DNA PRI 20-FEB-1999

VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

FEATURES  
 source

GC\_signal

GC\_signal

gene

5'UTR

exon

CDS

intron

exon

intron

BASE COUNT 954 a 1729 c 1677 g 1131 t  
 ORIGIN

Query Match 12.5% Score 437.4; DB 9; Length 5491;  
 Best Local Similarity 59.6% Pred. No. 1.7e-106;  
 Matches 1038; Conservative 0; Mismatches 506; Indels 197; Gaps 10;

QY 23 CCCGCCCTTGAGCAATGAGCCCGCGCTCTCGTTGCCCGCGGTGCGTCTCTGTCGG 82

AB016767.1 GI:4239869  
 telomerase transcriptase; hTERT.  
 Homo sapiens DNA.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
 Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 5491)  
 Takakura,M., Kyo,S., Kanaya,T., Takeda,J. and Inoue,M.  
 Direct Submission  
 Submitted (04-AUG-1998) to the DDBJ/EMBL/GenBank databases.  
 Masahiro Takakura, Kanazawa University, School of Medicine,  
 Department of Obstetrics and Gynecology; 13-1, Takaramachi,  
 Kanazawa, Ishikawa 920-0934, Japan  
 (E-mail:takakura@med.kanazawa-u.ac.jp, Tel:81-76-265-2425,  
 Fax:81-76-234-4266)  
 2 (sites)  
 Takakura,M., Kyo,S., Kanaya,T., Hirano,H., Takeda,J., Yutsudo,M.  
 Cloning and characterization of human telomerase caryotic subunit  
 (hTERT) gene promoter  
 Cancer Res. (1999) In press  
 Location/Qualifiers  
 1..5491  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 3250..3255  
 /note="telomerase transcriptase promoter"  
 3331..3336  
 /note="telomerase transcriptase promoter"  
 3338..5491  
 /gene="hTERT"  
 3338..3414  
 /gene="hTERT"  
 /evidence="experimental"  
 3338..3633  
 /gene="hTERT"  
 /number=1  
 join(3415..3633,3738..>5088)  
 /gene="hTERT"  
 /codon\_start=1  
 /product="telomerase transcriptase"  
 /protein\_id="BAA74724.1"  
 /db\_xref="PID:d1038457"  
 /db\_xref="PID:g4239870"  
 /db\_xref="GI:4239870"  
 /translation="MPAPRCRAVSLRSHREVLPATFVRRLPGQWRVQGGDP  
 AAFRALVAOLVCPMDARPPAPSPFQVSKELVARLQRLCERKAKNVLAFGA  
 LLDGARGPPEAFTTSVRSILPNTVDALRGSGAMGLLRVGDVLLVHLLARCLFV  
 LVNPSCAVOVCGPLYQLGAATQAPPPHAGSPRRRLCERANHSREAGVPLGLPA  
 PNARRGGASRLPLKPRRGAPEPEPTVGGGWAHPGTSGPSDRGFCVVSFA  
 RPAEATSLGALSGTRHSHPSVGRHAGPSTSRPRPHDTFCPPVIAETKRLFLYS  
 SDGKQLRPSFLLSLRLPLTGARLVETIFLGSPPCGQLPACGACATGKGCPC  
 FWSCLGTTRSAPTGCSRRRTARCELRSPOQVSPVSPRPMRPPTTQTPVAMCS  
 CSASTAPLAGVRLRAGLPAPAGAPRLQAOQRRRLNTKFTLSLGHAKLSLOELT  
 WKMSYRDCAWLRRSP"  
 3634..3737  
 /gene="hTERT"  
 /number=1  
 3738..5088  
 /gene="hTERT"  
 /number=2  
 5089..>5491  
 /gene="hTERT"  
 /number=2



CDS

/cell\_line="32D.3"

&lt;1..&gt;201

/codon\_start=1

/product="telomerase catalytic subunit"

/protein\_id="AAB84200.1"

/db\_xref="pid:g2605903"

/db\_xref="gi:2605903"

/translation="FFVYESTFQKNRFFYKSVKSLQSIGVROHLERVLRELSD

EVRHHDDTWLWPNICLRIFPK"

51 a 54 c 55 g 41 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 5.6%; Score 194.6; DB 12; Length 201;

Matches 197; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1686 TTCTTTTACATCAGAGACATCCAGAGACAGGCTCTCTTCTACCGTAAGAGT 1745

Db 1 TTCTTTTATGTCACGAGAGACATCCAGAGACAGGCTCTCTTCTACCGTAAGAGT 60

Qy 1746 GTGTGAGCAAGCTGAGAGATGGAGTACGAGCAACCTTGAGAGATGGGCTACGG 1805

Db 61 GTGTGAGCAAGCTGAGAGATGGAGTACGAGCAACCTTGAGAGATGGGCTACGG 120

Qy 1806 GAGCTGTACACAGAGAGGTCAGGCATCACCAGACACCTGGCTAGCCATGCCATCTGC 1865

Db 121 GAGCTGTACACAGAGAGGTCAGGCATCACCAGACACCTGGCTAGCCATGCCATCTGC 180

Qy 1866 AGACTGCGCTTCATCCCAAG 1886

Db 181 AGACTGCGCTTCATCCCAAG 201

RESULT 12

AF097365

LOCUS

DEFINITION

Homo sapiens

4356 bp DNA

and partial cds.

PRT

03-FEB-1999

ACCESSION

AF097365

NID

94210970

VERSION

AF097365.1

KEYWORDS

GI:4210970

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 4356)

Cong.F.S., Wen,J. and Bacchetti,S.

The human telomerase catalytic subunit hTERT: organization of the

gene and characterization of the promoter

Hum. Mol. Genet. 8 (1), 137-142 (1999)

99105927

REFERENCE

2 (bases 1 to 4356)

Cong,Y.S., Wen,J. and Bacchetti,S.

Direct Submission

Submitted (03-OCT-1998)

Pathology, McMaster University, 1200 Main

St. W., Hamilton, ON L8N 3Z5, Canada

Location/Qualifiers

1..4356

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

1..&gt;4356

/gene="TERT"

1..3996

/gene="TERT"

Join(3997..4215,4320..&gt;4356)

/gene="TERT"

/product="telomerase reverse transcriptase"

Join(3997..4215,4320..&gt;4356)

/gene="TERT"

/codon\_start=1

/product="telomerase reverse transcriptase"

/protein\_id="AAB12057.1"

/db\_xref="pid:94210971"

/db\_xref="gi:4210971"

/translation="MPRAPRCRAVRSLLRSHYREVLPLATFVRRLLGPGWRLVQRGDP

AFRALVACLVCPWDARPPPAAPSPFQVSCLELVARVL"

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 4.4%; Score 154.8; DB 11; Length 4356;

Matches 186; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Qy 23 CCGGGCTTTGACACAAATGACCCGCTCTCTGTCGCCGGGTCGCTCTCTCTCTCG 82

Db 3981 CCGGGCTTTGACACAAATGACCCGCTCTCTGTCGCCGGGTCGCTCTCTCTCG 82

Qy 83 CAGCGATACCGGAGGTGGCGCTGGCAACCTTTGTCGGCGCCTGGGCCCCGAGG 142

Db 4041 CAGCGATACCGGAGGTGGCGCTGGCAACCTTTGTCGGCGCCTGGGCCCCGAGG 142

Qy 143 CAGCGCTTTGTCAACCGGGGACCCGAGATCTTACCGCCTTTGTCGCCCAATGCC 202

Db 4101 CAGCGCTTTGTCAACCGGGGACCCGAGATCTTACCGCCTTTGTCGCCCAATGCC 202

Qy 203 AGTGTGATGCTGCGGCTACAGCTTCCACCTGCGGACCTTTCCTTCCACAGGTG 4160

Db 4161 GGTGTGCTGCTGGAGCGACGCGCGCCGCCCTCTCTTCCGCGAGGTG 4218

RESULT 13

AF098956

LOCUS

DEFINITION

Homo sapiens

2043 bp DNA

region and partial cds.

PRT

06-FEB-1999

ACCESSION

AF098956

NID

94226057

VERSION

AF098956.1

KEYWORDS

GI:4226057

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2043)

Horikawa,I., Cable,P.L., Afshari,C. and Barrett,J.C.

Cloning and characterization of the promoter region of human

telomerase reverse transcriptase gene

Cancer Res. 59 (4), 826-830 (1999)

99151529

REFERENCE

2 (bases 1 to 2043)

Horikawa,I., Cable,P.L., Afshari,C. and Barrett,J.C.

Direct Submission

Submitted (15-OCT-1998)

Laboratory of Molecular Carcinogenesis,

National Institute of Environmental Health Sciences, 111 T. W.

Alexander Drive, P.O. Box 12233, Research Triangle Park, NC 27709,

USA

Location/Qualifiers

1..2043

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="5"

/map="5p15.33"

1..&gt;2043

/gene="TERT"

1..1665

/gene="TERT"

1666..&gt;1939

/gene="TERT"

/product="telomerase reverse transcriptase"

1721..&gt;1939

/gene="TERT"

/note="catalytic subunit"



Db 169 TCCACTACAAGCCATATACCCGAATTCAGCACACTGGCGCGTTACTAGTGGATCC 110

Qy 3479 GAGCTCGGTACCAAGCTT 3496  
|||||

Db 109 GAGCTCGGTACCAAGCTT 92

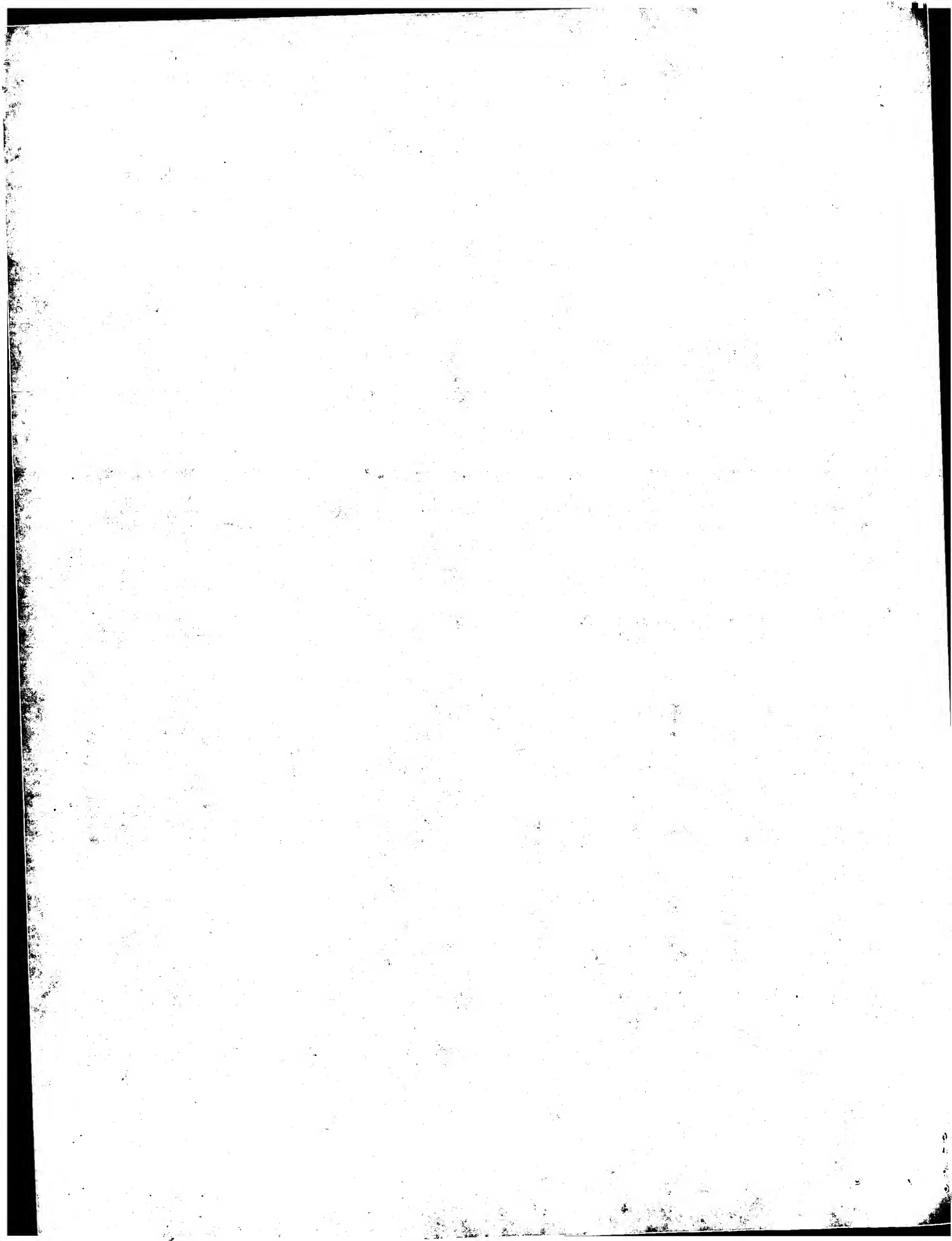
Search completed: August 13, 1999, 22:24:12  
Job time: 1799 sec

RESULT	15
HSP58NKRRC/c	
LOCUS	HSP58NKRRC 3418 bp DNA
DEFINITION	Homo sapiens DNA for p58 NK receptor gene.
ACCESSION	AJ000542
NID	G3282041
VERSION	AJ000542.1 GI:3282041
KEYWORDS	NK receptor; p58 gene.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	
Primates; Catarrhini; Hominoidea; Homo.	
Bianconi, R.	
Direct Submission	
Submitted (28-JUL-1997)	Bianconi R., Molecular Immunobiology, IST CBA, Largo R. Benzi 10, 16132, ITALY
2 (bases 1 to 3418)	
Bianconi, R.	
Genomic organization of p58 genes	
Unpublished	
Location/Qualifiers	
1..3418	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/chromosome="19"	
/cell_type="NK cells"	
762..3418	
/gene="p58"	
join(<762..797,3257..>3418)	
/gene="p58"	
/codon_start=3	
/product="NK receptor"	
/protein_id="CAA04175.1"	
/db_xref="PID:e1308408"	
/db_xref="PID:g3282042"	
/db_xref="GI:3282042"	
/db_xref="SPTREMBL:O75755"	
/translation="PFLLOGAVPHGVHRKPSLLAHPGLVKSSEVTILQCWSVMFE HFLLIHREGFKFDTLHLIKPNS"	
762..797	
/gene="p58"	
/number=2	
798..3256	
/gene="p58"	
/number=2	
3257..3418	
/gene="p58"	
/number=3	
BASE COUNT	853 a 765 c 1022 g 768 t 10 others
ORIGIN	
Query Match	1.8%; Score 63.6; DB 10; Length 3418;
Best Local Similarity	88.5%; Pred. No. 1.9e-06;
Matches	69; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY	3419 TCCGCTAGATGAACATGAAGGCCGAATTCCAGCACACTGGCGCGCTTACTAGTGATCC 3478

```

Query Match      1.8%  Score 63.6;  DB 10;  Length 3418;
Best Local Similarity 88.5%;  Pred. No. 1.9e-06;
Matches 69;  Conservative 0;  Mismatches 9;  Indels 0;  Gaps 0;
QQY 3419  TCCGCTAGCATCAACATGAAGGCCAAATCCAGACACTGGCGGCCGTACTAGTGATCC 3478
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```





GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 1999, 21:51:04 ; Search time 534.86 seconds

(without alignments)  
12893.051 Million cell updates/sec

Title: US-09-042-460-1  
Perfect score: 3496  
Sequence: 1 GAATCCGGGTGGGAGGCC.....CCGAGCTCGTACCAAGCTT 3496

Scoring table: IDENTITY\_NUC

Searched: 2546578 seqs, 986266752 residues

Database: EST:\*

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: em\_est10:\*
- 11: em\_est11:\*
- 12: em\_est12:\*
- 13: em\_est13:\*
- 14: em\_est14:\*
- 15: em\_est15:\*
- 16: em\_est16:\*
- 17: em\_est17:\*
- 18: em\_est18:\*
- 19: em\_est19:\*
- 20: gb\_est1:\*
- 21: gb\_est2:\*
- 22: gb\_est3:\*
- 23: gb\_est4:\*
- 24: gb\_est5:\*
- 25: gb\_est6:\*
- 26: gb\_est7:\*
- 27: gb\_est8:\*
- 28: gb\_est9:\*
- 29: gb\_est10:\*
- 30: gb\_est11:\*
- 31: gb\_est12:\*
- 32: gb\_est13:\*
- 33: gb\_est14:\*
- 34: gb\_est15:\*
- 35: gb\_est16:\*
- 36: gb\_est17:\*
- 37: gb\_est18:\*
- 38: gb\_est19:\*
- 39: gb\_est20:\*
- 40: gb\_est21:\*
- 41: gb\_est22:\*
- 42: gb\_est23:\*
- 43: gb\_est24:\*
- 44: gb\_est25:\*
- 45: gb\_est26:\*
- 46: gb\_est27:\*
- 47: gb\_est28:\*
- 48: gb\_est29:\*
- 49: gb\_est30:\*
- 50: gb\_est31:\*
- 51: gb\_est32:\*
- 52: em\_est20:\*
- 53: em\_est21:\*

- 54: em\_est22:\*
- 55: em\_est23:\*
- 56: em\_est24:\*
- 57: em\_est25:\*
- 58: em\_est26:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	235.6	6.7	389	31	AA281296
2	157.8	4.5	499	31	AA311750
3	91	2.6	375	30	AA200728
4	91	2.6	503	49	AI645957
5	57.8	1.7	140	22	R82515
6	56.8	1.6	209	40	AF027363
7	56.8	1.6	450	49	AF092755
8	56.2	1.6	503	34	AA514190
9	56.2	1.6	386	34	AA514191
10	55.4	1.6	281	49	AF121051
11	53.6	1.5	418	26	W91597
12	52.8	1.5	365	49	AF092809
13	52	1.5	295	41	AF064739
14	50.6	1.4	266	41	AF064731
15	50.4	1.4	239	22	U24210
16	45.2	1.3	555	26	W91680
17	44.4	1.3	197	49	AJ133823
18	42.8	1.2	474	30	AA262393
19	41.2	1.2	469	33	AA449622
20	39	1.1	397	33	AA401288
21	39	1.1	480	33	AA401288
22	39	1.1	377	34	AA504397
23	38.2	1.1	399	22	R72912
24	38.2	1.1	425	39	AA856459
25	38.2	1.1	436	47	AI509561
26	38	1.1	614	45	U42751
27	37.6	1.1	429	22	R22688
28	37.4	1.1	302	48	AI549853
29	37	1.1	502	42	AF074090
30	36.6	1.0	487	41	AA596393
31	36	1.0	470	39	AA859312
32	36	1.0	231	40	AA955313
33	36	1.0	582	43	AI177506
34	36	1.0	313	43	AI236238
35	36	1.0	456	46	AA925415
36	36	1.0	558	46	AI408893
37	35	1.0	422	51	AI705327
38	35	1.0	394	51	AI711279
39	35.8	1.0	278	39	AF077645
40	35.6	1.0	783	37	AA684336
41	35.6	1.0	436	44	AI269791
42	35.6	1.0	246	46	AI262622
43	35.6	1.0	512	50	AU067500
44	35.4	1.0	349	44	AI298513
45	35.2	1.0	279	23	H54560

ALIGNMENTS

RESULT 1

AA281296  
LOCUS  
DEFINITION  
zt08g02.r1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:712562 5',  
mRNA sequence.  
ACCESSION  
AA281296  
NID  
91924194  
VERSION  
AA281296.1 GI:1924194

AA281296 389 bp  
zt08g02.r1 NCI\_CGAP\_GCB1 Homo sapiens cDNA clone IMAGE:712562 5',  
mRNA sequence.  
ACCESSION  
AA281296  
NID  
91924194  
VERSION  
AA281296.1 GI:1924194

14-AUG-1997

us-09-042-460-1.1st

Mon Aug 16 10:21:26 1999

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
1 (bases 1 to 389)  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
JOURNAL On Sep 12, 1996 this sequence version replaced gi:1401009.  
COMMENT

Contact: Robert Strausberg, Ph.D.  
Tel: (301) 496-1550  
Email: Robert.Strausberg@nih.gov  
This clone is available royalty-free through LNL: contact the  
IMAGE Consortium (info@image.lni.gov) for further information.  
Insert Length: 2187 Std Error: 0.00  
Seq primer: -28m13 rev2 Et from Amersham  
High quality sequence stop: 385.

FEATURES  
source  
1..389  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:712562"  
/clone\_lib="NCI CGAP GCB1"  
/tissue\_type="germinal center B cell"  
/lab\_host="DH10B"  
/note="Vector: p7T3p-Pac (Pharmacia) with a modified  
polylinker. Site: 1: Not I; Site: 2: Eco RI; 1st strand cDNA  
was prepared from human tonsillar cells enriched for  
germinal center B cells by flow sorting (CD20+, IgD+),  
provided by Dr. Louis M. Staudt (NCI). Dr. David Allman  
(NCI) and Dr. Gerald Marti (CSER). cDNA synthesis was  
primed with a Not I - oligo(dT) primer  
[5'-GTGTTACCAATGAGTGGAGCGCGCCCTCATTTTTTTTTTTTTTTT-  
3']. Double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified p7T3 vector. Library  
went through one round of normalization, and was  
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 87 a 102 c 123 g 77 t  
ORIGIN  
Query Match 6.7%; Score 235.6; DB 31; Length 389;  
Best Local Similarity 75.6%; Pred. No. 2e-56; Indels 0; Gaps 0;  
Matches 292; Conservative  
QY 1632 GCTAGCTTCCTGTTCTGGCTGATGGACACATACGTTGGTACAGCTGCTTAGTCTATCTTTT 1691  
Db 1 GCCTAGTTCTCTGCACTGCTGATGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTTCTT 60  
QY 1692 TACATCAGAGAGAGACATCCAGAGACAGCTCTTCTTACCGTAAGAGTGTGG 1751  
Db 1 TACATCAGAGAGAGACATCCAGAGACAGCTCTTCTTACCGTAAGAGTGTGG 1751  
QY 1692 TACATCAGAGAGAGACATCCAGAGACAGCTCTTCTTACCGTAAGAGTGTGG 120  
Db 61 TATGTCAGGAGCACCCTTTCAAGAGACAGGCTTTTCTTACCGTAAGAGTGTGG 1811  
QY 1752 AGCAAGCTGCAGAGATTGGATGATGAGCAACCTTGTAGAGATGCGGTACGGGAGCTG 1811  
Db 1752 AGCAAGCTGCAGAGATTGGATGATGAGCAACCTTGTAGAGATGCGGTACGGGAGCTG 180  
QY 121 AGCAAGTTGCAAGACATTCGATCATCAGACAGCACTTGAAGAGGGTGCAGTGGGAGCTG 180  
Db 121 AGCAAGTTGCAAGACATTCGATCATCAGACAGCACTTGAAGAGGGTGCAGTGGGAGCTG 1871  
QY 1812 TCACAAGAGAGGAGTCTAGGATCATCAGGAGACCTTGGCTTAGCCATGCCATCTCCAGACTG 240  
Db 181 TCACAAGAGAGGAGTCTAGGATCATCAGGAGACCTTGGCTTAGCCATGCCATCTCCAGACTG 240  
QY 1872 CGCTTCATCCCAAGCCCAAGCGGCTCGGCGCCATTTGTAACATGAGTTATAGCATGGGT 1931  
Db 1872 CGCTTCATCCCAAGCCCAAGCGGCTCGGCGCCATTTGTAACATGAGTTATAGCATGGGT 300  
QY 241 CGCTTCATCCCAAGCCCTGAGCGGCTCGGCGCCATTTGTAACATGAGTTATAGCATGGGT 1991  
Db 1932 ACAGAGCTTTGGGAG 360  
QY 301 GCACAGAACGTTCCCGCAG 360  
Db 301 GCACAGAACGTTCCCGCAG 360

QY 1992 TTCAGCATGCTCACTACATGAGCGGAC 2017  
Db 361 TTCAGCGTCTCACTACATGAGCGGAC 386  
RESULT 2  
LOCUS AA311750 409 bp mRNA EST 19-APR-1997  
DEFINITION EST182469 Jurkat T-Cells VI Homo sapiens CDNA 5' end, mRNA  
sequence.  
AA311750  
ACCESSION AA311750  
NID G1964077  
VERSION AA311750.1 GI:1964077  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
AUTHORS Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 409)  
Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,  
Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,  
White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man, Wai, C.,  
Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,  
Fitzgerald, L.M., Fitzhugh, W.M., Fritchman, J.L., Geoghegan, N.S.,  
Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle, P.S., J.,  
Kelle, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,  
Moreno-Palancas, R.F., McDonald, D.A., Nguyen, D.T., Pelligrino, S.M.,  
Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R.,  
Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,  
Bednarek, D.P., Caci, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,  
Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A.,  
He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,  
Kozak, D.L., Kunsch, C., Hung, J., Li, H., Weissner, P., Olsen, H.,  
Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M.,  
Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,  
Fraser, C.M., and Venter, J.C.  
Initial assessment of human gene diversity and expression patterns  
based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)  
JOURNAL 96026280  
MEDLINE  
COMMENT On Sep 12, 1996 this sequence version replaced gi:1402064.

Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org  
For clone availability, additional sequence and expression  
information related to this EST, please check the TIGR Human Gene  
Index (<http://www.tigr.org/tdb/hgi/hgi.html>)  
Seq primer: M13 Reverse.

FEATURES  
source  
1..409  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="Jurkat T-cells VI"  
/cell\_type="T-lymphocyte"  
/note="Vector: pBluescript SK-; Site\_1: EcoRI; Site\_2:  
XbaI"

BASE COUNT 65 a 120 c 133 g 86 t 5 others  
ORIGIN  
Query Match 4.5%; Score 157.8; DB 31; Length 409;  
Best Local Similarity 75.3%; Pred. No. 3e-34;  
Matches 195; Conservative  
QY 2611 GTTTTGTGATGACTTCTTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2670  
Db 1 GTTTTGTGATGACTTCTTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 60





DEFINITION AF092755 Toxoplasma gondii encystation, subtracted cDNA Toxoplasma gondii cDNA clone CL17, mRNA sequence.

ACCESSION AF092755

NID 94731799

VERSION AF092755.1 GI:4731799

KEYWORDS EST.

SOURCE Toxoplasma gondii.

ORGANISM Toxoplasma gondii.

REFERENCE Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae; Toxoplasma.

AUTHORS 1 (bases 1 to 450)

TITLE Yahiaoui,B., Dzielerszinski,F., Bernigaud,A., Slomianny,C., Camus,D. and Tomavo,S.

JOURNAL Isolation and characterization of a subtractive library enriched for developmentally regulated transcripts expressed during encystation of Toxoplasma gondii

COMMENT Mol. Biochem. Parasitol. 99, 223-235 (1999)

On Mar 10, 1998 this sequence version replaced gi:2948993.

Contact: Stanislas Tomavo  
Laboratoire de Chimie Biologique, CNRS UMR 8576  
Batiment C9, Universite des Sciences et Technologies de Lille,  
59655 Villeneuve d'Ascq, France  
Email: Stan.Tomavo@univ-lille.fr.

FEATURES source

Location/Qualifiers

1..450

/organism="Toxoplasma gondii"

/db\_xref="taxon:5811"

/clone="CL17"

/clone\_lib="Toxoplasma gondii encystation, subtracted cDNA"

/dev\_stage="encystation"

/note="Vector: T/A cloning vector pCRII(Invitrogen)"

BASE COUNT 92 a 108 c 142 g 105 t 3 Others

ORIGIN

Query Match 1.6%; Score 56.8; DB 49; Length 450;

Best Local Similarity 96.7%; Pred. No. 1.8e-05;

Matches 58; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3437 AGGCGCAATTCACGACACTGGCGCGTTACTAGTGGATCGGACCTCGGTACCAAGCTT 3496

Db 65 AAGCCGAATTCACGACACTGGCGCGTTACTAGTGGATCGGACCTCGGTACCAAGCTT 6

RESULT 8

AA514190/c

LOCUS AA514190 503 bp mRNA EST 14-JUL-1997

DEFINITION HFLEST-741 Human fetal liver (S.Xue) Homo sapiens cDNA, mRNA

sequence.

ACCESSION AA514190

NID 92253714

VERSION AA514190.1 GI:2253714

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Du,Q., He,F. and Xue,S.

TITLE Human liver ESTs

JOURNAL Unpublished (1997)

COMMENT On May 18, 1995 this sequence version replaced gi:811227.

Contact: Shepu Xue  
Dept of Cellular Biology  
Institute of Basic Medical sciences/Peking Union Medical College &  
Chinese Academy of Medical Sciences(PUMC & CAMS)  
5 Dong Dan San Tiao, Beijing, 100005 P.R. China  
Tel: 8601-65296459  
Fax: 8610-65240529  
Email: Xuesp@cdm.imicams.ac.cn

FEATURES source

Location/Qualifiers

1..503

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="4; 14"

/clone\_lib="Human fetal liver (S.Xue)"

/tissue\_type="liver"

/dev\_stage="fetal"

/lab\_host="E.coli DH5a"

/note="vector: pBluescript SK"

BASE COUNT 128 a 124 c 118 g 133 t

ORIGIN

Query Match 1.6%; Score 56.2; DB 34; Length 503;

Best Local Similarity 95.1%; Pred. No. 2.9e-05;

Matches 58; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3436 AAGCGCAATTCACGACACTGGCGCGTTACTAGTGGATCGGACCTCGGTACCAAGCT 3495

Db 67 ATGACGAATTCACGACACTGGCGCGTTACTAGTGGATCGGACCTCGGTACCAAGCT 8

QY 3496 T 3496

Db 7 T 7

RESULT 9

AA514191

LOCUS AA514191 386 bp mRNA EST 14-JUL-1997

DEFINITION HFLEST-742 Human fetal liver (S.Xue) Homo sapiens cDNA, mRNA

sequence.

ACCESSION AA514191

NID 92253715

VERSION AA514191.1 GI:2253715

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Du,Q., He,F. and Xue,S.

TITLE Human liver ESTs

JOURNAL Unpublished (1997)

COMMENT On May 18, 1995 this sequence version replaced gi:811228.

Contact: Shepu Xue  
Dept of Cellular Biology  
Institute of Basic Medical sciences/Peking Union Medical College &  
Chinese Academy of Medical Sciences(PUMC & CAMS)  
5 Dong Dan San Tiao, Beijing, 100005 P.R. China  
Tel: 8601-65296459  
Fax: 8610-65240529  
Email: Xuesp@cdm.imicams.ac.cn

Seq primer: M13 Forward and Reverse Primer.

Location/Qualifiers

1..386

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="14"

/clone\_lib="Human fetal liver (S.Xue)"

/tissue\_type="liver"

/dev\_stage="fetal"

/lab\_host="E.coli DH5a"

/note="vector: pBluescript SK"

BASE COUNT 90 a 97 c 100 g 99 t

ORIGIN

Query Match 1.6%; Score 56.2; DB 34; Length 386;

Best Local Similarity 95.1%; Pred. No. 2.5e-05;

Matches 58; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Mon Aug 16 10:21:26 1999

QY 3436 AAGGCGCAATCCACACACTGGCGCGCTTACTAGTGGATCCGAGCTCGGTACCAAGCT 3495  
 Db 118 ATGGAGCAATCCACACACTGGCGCGCTTACTAGTGGATCCGAGCTCGGTACCAAGCT 177  
 QY 3496 T 3496  
 Db 178 T 178  
 RESULT 10  
 AF121051/c 281 bp mRNA EST 21-APR-1999  
 LOCUS AF121051 Homo sapiens liver fetal Homo sapiens cDNA clone TTG-L8,  
 DEFINITION mRNA sequence.  
 ACCESSION AF121051  
 NID 94630719  
 VERSION AF121051.1 GI:4630719  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1. (bases 1 to 281)  
 AUTHORS Zhu, T., Zhang, L., Liu, Z., and Zhang, J.  
 TITLE Homo sapiens chromosome 11p13 EST  
 JOURNAL Unpublished (1999)  
 COMMENT On Mar 20, 1998 this sequence version replaced gi:2980587.  
 Contact: Zhu T  
 Medical College  
 Nankai University  
 94 Weijin Rd., Tianjin, 300071, People's Republic of China.  
 FEATURES source  
 Location/Qualifiers  
 1. 281  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /map="3: 11p13"  
 /clone="TTG-L8"  
 /clone\_lib="Homo sapiens liver fetal"  
 /tissue\_type="liver"  
 /dev\_stage="fetal"  
 /dev\_stage="fetal"  
 BASE COUNT 61 a 65 c 101 g 54 t  
 ORIGIN  
 Query Match 1.6%; Score 55.4; DB 49; Length 281;  
 Best Local Similarity 90.8%; Pred. No. 3.4e-05;  
 Matches 59; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
 QY 3432 CATGAAGCGCAATCCACACACTGGCGCGCTTACTAGTGGATCCGAGCTCGGTACCA 3491  
 Db 65 CACAGCGCAATCCACACACTGGCGCGCTTACTAGTGGATCCGAGCTCGGTACCA 6  
 QY 3492 AGCTT 3496  
 Db 5 AGCTT 1  
 RESULT 11  
 W91597 418 bp mRNA EST 09-JUL-1996  
 LOCUS W91597  
 DEFINITION MTA.G05.077.A MTA adult mouse thymus library Mus musculus cDNA  
 clone MTA.G05.077 3' end similar to CT repeat, mRNA sequence.  
 ACCESSION W91597  
 NID g1408023  
 VERSION W91597.1 GI:1408023  
 KEYWORDS EST.  
 SOURCE house mouse.  
 ORGANISM Mus musculus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 418)

AUTHORS Nguyen, C., Rocha, D., Granjeaud, S., Bernard, K., Naquet, P. and Jordan, B.R.  
 TITLE Gene expression in different cell types of the mouse thymus  
 JOURNAL Unpublished (1996)  
 COMMENT On Apr 14, 1993 this sequence version replaced gi:785863.  
 Contact: Jordan BR  
 Genome Structure and Immune Functions  
 Centre d'Immunologie INSERM/CNRS  
 Case 906, 13288 MARSEILLE Cedex 9, FRANCE  
 Tel: 330491269496  
 Fax: 330491269430  
 Email: jordan@ciml.univ-mrs.fr  
 This sequence was determined at Engelhardt Institute of Molecular  
 Biology, Moscow, Russia (V Zakhariev)  
 Seq primer: SP6.  
 FEATURES Location/Qualifiers  
 1. 418  
 /organism="Mus musculus"  
 /strain="C57Bl/6"  
 /db\_xref="taxon:10090"  
 /map="17q21"  
 /clone="MTA.G05.077"  
 /clone\_lib="MTA adult mouse thymus library"  
 /lab\_host="MC1061 p3"  
 /note="Vector: pcDNA1; Site\_1: NotI; Site\_2: EcoRI; The  
 cDNA library was constructed from poly(A)+ RNA of an adult  
 mouse thymus by oligo-dT primed reverse transcription.  
 cDNA was selected on gel for size above 800 nucleotides  
 after second strand synthesis, then directionally cloned  
 into the pcDNA1 vector (Not I on polyA side, EcoR I on the  
 5' side)."  
 BASE COUNT 111 a 102 c 85 g 114 t 6 others  
 ORIGIN  
 Query Match 1.5%; Score 53.6; DB 26; Length 418;  
 Best Local Similarity 81.6%; Pred. No. 0.00014;  
 Matches 62; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
 QY 3421 CCTAGATGACATCAAGGGCGAATCCAGCACACTGGCGCGCTTACTAGTGGATCCGA 3480  
 Db 265 CTCAAAAAAGAGACGTCTGTAATCCACACACTGGCGCGCTTACTAGTGGATCCGA 324  
 QY 3481 GCTCGGTACCAAGCTT 3496  
 Db 325 GCTCGGTACCAAGCTT 340  
 RESULT 12  
 AF092809/c 365 bp mRNA EST 03-MAY-1999  
 LOCUS AF092809 Toxoplasma gondii encystation, subtracted cDNA Toxoplasma  
 DEFINITION gondii cDNA clone CL146, mRNA sequence.  
 ACCESSION AF092809  
 NID 94731853  
 VERSION AF092809.1 GI:4731853  
 KEYWORDS EST.  
 SOURCE Toxoplasma gondii.  
 ORGANISM Toxoplasma gondii  
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;  
 Sarcocystidae; Toxoplasma.  
 REFERENCE 1 (bases 1 to 365)  
 AUTHORS Yakhiaoui, B., Dzierszinski, F., Bernigaud, A., Slomiany, C., Camus, D.  
 TITLE Isolation and characterization of a subtractive library enriched  
 for developmentally regulated transcripts expressed during  
 encystation of Toxoplasma gondii  
 JOURNAL Mol. Biochem. Parasitol. 99, 223-235 (1999)  
 COMMENT On Mar 10, 1998 this sequence version replaced gi:2949046.  
 Contact: Stanislas Tomavo



Laboratoire de Chimie Biologique, CNRS UMR 8576  
Batiment C9, Université des Sciences et Technologies de Lille,  
59655 Villeneuve d'Assq, France  
Email: Stan.Tomavo@univ-lille1.fr.

FEATURES  
source  
1. .365  
/organism="Toxoplasma gondii"  
/db\_xref="taxon:5811"  
/map="5"  
/clone="CL146"  
/clone\_lib="Toxoplasma gondii encystation, subtracted  
cDNA"  
/dev\_stage="encystation"  
/note="vector: T/A cloning vector pCRII(Invitrogen)"  
73 a 91 C 112 g 80 t 9 others  
BASE COUNT  
ORIGIN  
Query Match 1.58; Score 52.8; DB 49; Length 365;  
Best Local Similarity 93.18; Pred. No. 0.00022;  
Matches 54; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 3435 GAAGGGCAATTCAGCACACTGGCGCGCTTACTAGTGGATCGGCTCGGTACCAA 3492  
|||||  
Db 65 GAAGGGCAATTCAGCACACTGGCGCGCTTACTAGTGGATCGGCTCGGTACCAA 8

RESULT 13  
AF064739  
LOCUS  
DEFINITION AF064739 295 bp mRNA EST 30-OCT-1998  
similar to protein tyrosine phosphatase, mRNA sequence.  
AF064739  
NID 93242808  
VERSION AF064739.1 GI:3242808  
KEYWORDS  
SOURCE clove pink.  
ORGANISM Diantthus caryophyllus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Caryophyllales; Caryophyllaceae;  
Diantthus.  
1 (bases 1 to 295)  
Kim, J.Y., Chung, Y.S. and Shin, J.S.  
Isolation of differentially expressed genes during flower  
development and pigmentation in carnation  
Unpublished (1998)  
On Jan 17, 1998 this sequence version replaced gi:2045312.  
Contact: Jeong Sheop Shin  
Plant Molecular Genetics  
Graduate School of Biotechnology, University of Korea  
136-701 Anam-dong 5/1 Seoul, Korea  
Tel: 00 82 2 3290 3430  
Fax: 00 82 2 927 9028  
Email: jsshine@kucn.korea.ac.kr.  
Location/Qualifiers  
1. .295  
/organism="Diantthus caryophyllus"  
/db\_xref="taxon:3570"  
/clone="cpi-10"  
/clone\_lib="Carnation mRNA"  
85 a 57 C 49 g 104 t  
BASE COUNT  
ORIGIN

Query Match 1.58; Score 52; DB 41; Length 295;  
Best Local Similarity 100.0%; Pred. No. 0.00033;  
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 3435 GAAGGGCAATTCAGCACACTGGCGCGCTTACTAGTGGATCGGCTCGG 3486  
|||||  
Db 243 GAAGGGCAATTCAGCACACTGGCGCGCTTACTAGTGGATCGGCTCGG 294

RESULT 14  
AF064731  
LOCUS  
DEFINITION AF064731 266 bp mRNA EST 30-OCT-1998  
similar to catalase, mRNA sequence.  
AF064731  
NID 93242800  
VERSION AF064731.1 GI:3242800  
KEYWORDS  
SOURCE EST.  
ORGANISM Diantthus caryophyllus  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core  
eudicots; Caryophyllales; Caryophyllaceae;  
Diantthus.  
1 (bases 1 to 266)  
Kim, J.Y., Chung, Y.S. and Shin, J.S.  
Isolation of differentially expressed genes during flower  
development and pigmentation in carnation  
Unpublished (1998)  
On Jan 17, 1998 this sequence version replaced gi:2045306.  
Contact: Jeong Sheop Shin  
Plant Molecular Genetics  
Graduate School of Biotechnology, University of Korea  
136-701 Anam-dong 5/1 Seoul, Korea  
Tel: 00 82 2 3290 3430  
Fax: 00 82 2 927 9028  
Email: jsshine@kucn.korea.ac.kr.  
Location/Qualifiers  
1. .266  
/organism="Diantthus caryophyllus"  
/db\_xref="taxon:3570"  
/map="2; 2; 2; 15"  
/clone="cpi-2"  
/clone\_lib="Carnation mRNA"  
70 a 69 C 58 g 69 t  
BASE COUNT  
ORIGIN

Query Match 1.4%; Score 50.6; DB 41; Length 266;  
Best Local Similarity 93.0%; Pred. No. 0.00077;  
Matches 53; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 3440 GCGAATTCAGCACACTGGCGCGCTTACTAGTGGATCGGCTCGGCTCGG 3496  
|||||  
Db 209 GCGAATTCAGCACACTGGCGCGCTTACTAGTGGATCGGCTCGGCTCGG 265

RESULT 15  
U24210  
LOCUS  
DEFINITION MMU24210 Mouse cerebellum Mus musculus cDNA clone PBS114, MRNA  
sequence.  
U24210  
NID 91002362  
VERSION U24210.1 GI:1002362  
KEYWORDS  
SOURCE EST.  
ORGANISM Mus musculus  
house mouse.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;  
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 239)  
Wei, J., Hodes, M., Wang, Y., Feng, Y., Chetti, B. and Dlouhy, S.  
Direct cDNA selection from hybridization of mouse cerebellar cDNA  
libraries against microdissection DNA pool in MMU16C3-C4 region  
Unpublished (1995)  
Contact: Wei, J.  
Medical Genetics  
IUPUI

us-09-042-460-1.rst

Mon Aug 16 10:21:26 1999

975 W. Walnut St., Indianapolis, IN 46202, USA

Insert Length: 1421 Std Error: 0.00

High quality sequence stop: 221.

# FEATURES

source

Location/Qualifiers

1. .239

/organism="Mus musculus"

/strain="B6CBA"

/db\_xref="taxon:10090"

/map="MMU16C3-C4 region; MMU16C3-C4 region"

/clone="PBS114"

/clone\_lib="Mouse cerebellum"

/note="Newborn (P1) mice."

66 a 55 c 59 g 59 t

BASE COUNT

ORIGIN

Query Match 1.4%; Score 50.4; DB 22; Length 239;

Best Local Similarity 90.0%; Pred. No. 0.00082;

Matches 54; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 3437 AGGGGAATTCACACACTGGCGCGTTACTAGTGGATCCGAGCTCGGTACCAAGCTT 3496

Db 180 AGCCGAATTCACACACTGGCGCGTTACTAGTGGATCCGAGCTCGGTACCAAGTT 239

Search completed: August 13, 1999, 22:05:56

Job time: 892 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 1999, 11:20:47 ; Search time 79.57 Seconds  
(without alignments)  
10992.479 Million cell updates/sec

Title: US-09-042-460-1  
Perfect score: 3496  
Sequence: 1 GAATTCGGGTGGGAGGCC.....CCGAGCTCGGTACCAAGCTT 3496

Scoring table: IDENTITY\_NUC

Searched: 311585 seqs, 125096042 residues

Database : N\_Geneseq\_36:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1600.8	45.8	3955	1 V22379	Human telomerase r
2	1600.8	45.8	4023	1 V60320	Human telomerase g
3	1600.8	45.8	4042	1 V72117	Human catalytic te
4	1599.2	45.7	3798	1 V27876	Human telomerase p
5	1594.4	45.6	4037	1 V22428	Human telomerase r
6	1323.2	37.8	3855	1 V22382	Human telomerase r
7	1284.6	36.2	2848	1 V27872	Human telomerase p
8	992.6	28.4	2357	1 X15923	CDNA encoding a ca
9	826.8	23.6	2089	1 V72126	Human catalytic te
10	715	20.5	2175	1 X15925	Nucleotide sequenc
11	714.4	20.4	2176	1 V22380	Human telomerase r
12	691.4	19.8	2171	1 V22426	Human telomerase r
13	647	18.5	3346	1 V60321	Human telomerase r
14	582.8	16.7	3972	1 V72125	Human telomerase r
15	468.8	13.4	4335	1 V16979	Human catalytic te
16	336.2	9.6	949	1 V27875	Human telomerase r
17	275.4	7.9	523	1 X15926	Nucleotide sequenc
18	235.6	6.7	395	1 X15924	Human catalytic te
19	228.8	6.5	1012	1 V72124	Human telomerase r
20	119.2	3.4	1153	1 V72122	Human catalytic te
21	80.8	2.3	182	1 V16980	Human telomerase r
22	60.8	1.7	1936	1 T35333	Peroxisome prolif
23	58.6	1.7	1242	1 T89896	Clam cyclin-select
24	58	1.7	9837	1 T96851	Intron 21 of human
25	56.8	1.6	8136	1 T63236	Plasmid pCR(RM)33a
26	56.8	1.6	6115	1 T63235	Eukaryotic express
27	56.8	1.6	1272	1 V19124	Nucleotide sequenc
28	55.4	1.6	5120	1 V01426	Plasmid pCMV-SmgST
29	55	1.6	3984	1 T07310	Vector plasmid p70
30	55	1.6	2178	1 T15929	Neural cell adhesi
31	55	1.6	9592	1 T36752	Adenovirus vector
32	55	1.6	5540	1 T62548	IgSP-hpOMdelACTH-
33	55	1.6	10443	1 T71319	Plasmid pPB72 enco
34	55	1.6	5446	1 V38297	Sequence of pGEM C
35	55	1.6	6420	1 X15625	Human Flt4 ligand
36	54.8	1.6	219	1 V48407	DNA molecule encod
37	54.6	1.6	908	1 V68843	Murine metastatic
38	54.4	1.6	404	1 T69381	Vector back-end ol
39	54	1.5	94	1 Q89696	Plasmid pRC/ASK-S
40	54	1.5	6216	1 V01425	Human rhodopsin cd
41	53.4	1.5	573	1 T70422	Wheat low affinity
42	53.4	1.5	1982	1 V10366	Nucleotide sequenc
43	53.4	1.5	4525	1 V69746	

Eukaryotic express  
Eukaryotic (CMV) e

## ALIGNMENTS

```
RESULT 1
V22379
ID V22379 standard; CDNA; 3955 BP.
AC V22379;
DT 13-AUG-1998 (first entry)
DE Human telomerase reverse transcriptase encoding cDNA
KW Human; telomerase reverse transcriptase; hTERT; TPT; diagnosis;
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 56..3454
FT FT /*tag= a
FT /product= "telomerase reverse transcriptase"
PN GB3317891-A.
PD 08-APR-1998.
PF 01-OCT-1997; 020890.
PR 14-AUG-1997; US-915503.
PR 01-OCT-1996; US-724643.
PR 18-APR-1997; US-844419.
PR 25-APR-1997; US-846017.
PR 06-MAY-1997; US-851843.
PR 09-MAY-1997; US-854050.
PR 14-AUG-1997; US-911312.
PR 14-AUG-1997; US-9112951.
PA (GEO-) Geron Corp.
PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,
PI Morin GB, Nakamura T, Harley CB;
PI WPI; 98-171633/16.
DR P-PDB; W46957.
PT Pure and recombinant human telomerase reverse transcriptase
PT variants - are useful in the diagnosis, prognosis and treatment of
PT cell proliferation conditions especially cancer and ageing
PS Claim 5; Fig 16; 387pp; English.
CC The present sequence encodes human telomerase reverse transcriptase
CC (hTERT), which is a ribonucleoprotein. The present invention also
CC describes the following methods: (A) determining whether a test compound
CC is a modulator of hTERT, by detecting the change in hTERT recombinant
CC protein or polynucleotide on administration of the compound; (B)
CC preparation of recombinant telomerase by contacting a protein
CC the hTERT RNA or protein in a sample by binding a relevant probe to the
CC sample and detecting the complex formed or in the case of RNA detection,
CC amplifying the product with presence of hTERT in the sample; and (D)
CC increasing the proliferation of a vertebrate cell by increasing hTERT
CC expression; and (E) the use of an agent that causes an increase in cell
CC ageing. A protein preparation of hTERT and the polynucleotide encoding
CC hTERT can be used in the manufacture of medicaments for inhibiting the
CC effect of ageing or cancer. Inhibitors of telomerase activity can be
CC used to treat conditions that are associated with high telomerase
CC activity. A protein preparation of hTERT can also be used in the new
CC methods.
SQ Sequence 3955 BP; 648 A; 1353 C; 1251 G; 703 T;
```

Query Match 45.8%; Score 1600.8; DB 1; Length 3955;  
Best Local Similarity 68.9%; Pred. No. 0;  
Matches 2379; Conservative

```
QY 23 CCGGGCTTCGACACATACCGCGCTCCTCGTGGCCCGCGTCTCTGCTGCG 82
      ||||| | | | | | | | | | | | | | | | | | | | | | | | |
Db 40 CCGGGCACCCCGCGCGCTCCCGCTCGCGAGCGCGTCTCTGCTGCG 99
      ||||| | | | | | | | | | | | | | | | | | | | | | | |
QY 83 CACCGATACCGGAGGTGTGGCGCTGGCAACCTTTGTGGCGCGCTGGCGCCGAGG 142
```

		1150	TCTGGTTCCAGGBCCTCGATGCCAGGACTCCCGCAGAGTTCGCCGCCTGCCCCAGCG	1209
Dd				
		1208	ATACATGGCAGATGCGGCCCTCTTTCCAAACAGCTGCTGTGAACCATCAGAGTGCCCAATA	1267
Qy				
		1210	CTATCGCAAATGCGGCCCTGTTTCTGGAGCTGCTTGGGAACACCAGCGAGTGCCCTTA	1269
Dd				
		1268	TGTCAGACTCTCAGTTCACATTGACGTTTTGAAACAGCAACAACAGTAGTGCAGATGC	1327
Qy				
		1270	CGGGTCCTCTCTCAAGACGACTGCCGCTGGAGCTCGGTCTCACCCCAGCAGCGGCTGT	1329
Dd				
		1328	CT-----TGAAACACAGGCC	1342
Qy				
		1330	CTGTGCCCCGGAGAAGCCCGAGGCTCTGTGCGGCCCCCGAGGAGGAGGACACAGACCC	1389
Dd				
		1343	ACGCACTCATGGATTGCTCGGCTGCACAGCAGTCCCTGGCAGGTATATGTTTTCT	1402
Qy				
		1390	CCGTGCGCTGGTCAGCTGCTCGCCAGCAGCAGACGCCCTGGCAGGTGTACGGCTCGT	1449
Dd				
		1403	TCGGGCTGTCTCTGAAAGTGTGTCTGCTAGTCTCTGGGTACCAGCACAATGAGCG	1462
Qy				
		1450	CGGCGCTGCTGCGCGGCTGCTGCCCCAGCGCTCTGGGCTCCAGGCACACGAACG	1509
Dd				
		1463	CCGCTCTTTAAGAACTTAAGAGATTCACTCTGTTGGGGAATAACGGCAAGCTATFAC	1522
Qy				
		1510	CCGCTTCTCAGGAACACCAAGAGTTTCACTCTCTGGGGAACATGCCAAGCTCTCGCT	1569
Dd				
		1523	GCAGCACTGATGTGGAAGATGAAAGTAGAGATTGCCATGCTCGCAGCAGCCCGG	1582
Qy				
		1570	GCGAGCTGACGTGGAAGATGAGCGTCCGAGACTGCGCTTGGCTGCGCAGGAGCCAGG	1629
Dd				
		1583	GAAGACCGTGTCCCGCTGCAGAGCAGCTGTCAGGAGAGGATCTGCTACGTTCT	1642
Qy				
		1630	GCTGGCTGTGTTCCGGCCGAGAGCAGCTGCTGCGTAGGAGATCTGCGCAAGTTCT	1689
Dd				
		1643	GTCTGGCTGATGGACACATACATGCTGTGACGCTCTTAGTTCATCTTTTACATCAGA	1702
Qy				
		1690	GCATGGCTGATGAGTGTACGTCGTGAGCTGCTCAGGCTCTTTTATGTCACGGA	1749
Dd				
		1703	GAGCAATTCAGAGAAGACAGGCTCTTCTTACCGTAAGAGTGTGTGGAGCAAGCTGCA	1762
Qy				
		1750	GACCACTTTCAAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTGCA	1809
Dd				
		1763	GAGCATGGAGTCAGCAACACCTTGAGAGTGGCGGTACGGAGCTGTCCACAGAGGA	1822
Qy				
		1810	AAGCATTGGATCAGACAGCACTTGAAGAGGCTGCAGCTGGGGAGCTGTCCGAAGACA	1869
Dd				
		1823	GGTCAGSCATCACAGGACACCTGCTAGCCATGCCATCTCAGACTGCGCTTCATCCC	1882
Qy				
		1870	GGTCAGCAGCATCGGGAAGCCAGGCCCCCTGCTGACGCTCCAGACTCCGCTTCATCCC	1929
Dd				
		1883	CAAGCCCAAGCGCTCGGCGCCATTGTGAACATGAGTTATAGCATGGGTACAGAGCTTT	1942
Qy				
		1930	CAAGCTTACGAGGCTGGCGGCTGGCGGCTGTGAACATGGACTACGCTGGGAGCAGCT	1989
Dd				
		1943	GGCAGAGAAGAACAGCGCCAGCATTTACCCAGCTCTCAGACTCTCTTCTCAGCATGCT	2002
Qy				
		1990	CCGAGAAAAAGAGGCGGAGCGTCTACCTCGAGGGTGAAGGCACTGTTTACGCTGCT	2049
Dd				
		2003	CAACTATGAGCGGACAAAACATCTCACCCTTATGGGCTCTTCTGTACTGGGTATGATGA	2062
Qy				
		2050	CAACTACAGCGGCGGCGGCGCCCGCCCTCCCTGGCGGCTCTGTGTGGGCTTGACGA	2109
Dd				
		2063	CATCTACAGGACCTGGCGGCGCTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2122
Qy				
		2110	TATCAGAGGCTGGCGACCTTCGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	2169
Dd				
		2123	GATGATCTTGTATAGGACAGATGTACCGGCGCTATGATGCCATCCCGGAGGTAAAGT	2182
Qy				
		2170	GCTGACTTGTTCAGAGGTGATGTACGCGGCGCTACGACACCATCCCGGAGCAGGCT	2229
Dd				
		2183	GGTGAGGTTGTGTGCAATATATGATCAGGCACTCGGAGAGCAGCTACTGTATCCCGCAGTA	2242
Qy				
		2230	CACGAGGTCATCGCCAGCATCATCA---AACCCCAACACGTAATCTGCTGCTGCTGCTA	.2286
Dd				

QY 2243 TGCAGTGTGTCGGAGAGATAGCCAAAGCCCAAGTCCCAAGTCTCTTAGGAGACAGGTAC 2302  
DB 2287 TGCAGTGTGTCAGAAAGCCGCCCTTGGGACAGTCCGCAAGGCTTCAAGAGCCAGCTCTC 2346  
QY 2303 CACCTCTCTGACCTCCAGCATTACATGGCCAGTTCCTTAAGCATCTCGAGATTTCAGA 2362  
DB 2347 TACCTTGACAGACTCCAGCGGTACATGCGACAGTTCGTGGGTCACTCTCAG-----GA 2400  
QY 2363 TGCAGTGCCTGAGGACTCCCTGTGTCATCGAGCAGCATCTCTATGAATGAGAGCAG 2422  
DB 2401 GACAGCCCTGAGGAGTCCCTGCTCATCGAGCAGAGCTCTCTCTGAATGAGGCGCAG 2460  
QY 2423 CAGCAGCCTCTTTGACTTCTCTGACACTTCTGCTGACATGCTGCTAAAGATTGGTGA 2482  
DB 2461 CAGTGGCTCTTTCAGAGCTCTCTGACGCTTCATGTGACACAGCGCTGCGCATCAGGG 2520  
QY 2483 CAGGTGCTATACAGTGCAGGGCATCCCGAGGGCTCCAGCCTATCACCTGCTCTG 2542  
DB 2521 CAAGTCTTACGTCAGTGCAGGGATCCGAGGGCTCCATCTCTCCAGCTGCTCTG 2580  
QY 2543 CAGTCTGTGTCGGAGACATGAGACAGCTGTTGCTGAGGTGACGGGATGGGT 2602  
DB 2581 CAGCTGTGCTAGCGGACATGAGAACAGCTGTTGCGGGATTCGGCGGACGGCT 2640  
QY 2603 GCTTTTACGTTTGTGATGACTTCTGTTGAGCGCTCCTGTTGAGACAGCTCCATACAGCT 2682  
DB 2641 GCTCTGCGTTTGTGGATGATTTCTGTTGGTGAACCTCCTCACCCACGCGAAGC 2700  
QY 2663 CTTCCTCAGCAGCTGTCATGCGGTTCTGAGTATGGTGCATGATAAATTCAGAA 2722  
DB 2701 CTTCCTCAGCAGCTGTCATGCGGTTCTGAGTATGGTGCATGAGTATGGTGCATG 2760  
QY 2723 GACAGTGTGAACTTCCCTGTCGAGCTGTTGAGTATGGTGCATGAGTATGGTGCATG 2782  
DB 2761 GACAGTGTGAACTTCCCTGTCGAGCTGTTGAGTATGGTGCATGAGTATGGTGCATG 2820  
QY 2783 GCCTGCTCAGCTGTTTCCCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2842  
DB 2821 GCGGCTTCAAGGCTGAGGAGAACATGCGTCAAACTCTTTGGGCTTTCGGCTGAGGT 2880  
QY 2843 GTTCTGTGACTTACCTGAGTATGCGGAGCTCAATTAAGCAGGCTCAGCTTCCAGAG 2902  
DB 2881 GCAGAGCAGCTTACCTGAGCTATGCGGAGCTCCATCAGAGCAGCTCAGCTTCAACG 2940  
QY 2903 TGTCTTCAAGCTGGGAGAACATGCGGAAACAGCTCTGCTGCTGCTGCTGCTGCTGCTG 2962  
DB 2941 CGGCTTCAAGGCTGAGGAGAACATGCGTCAAACTCTTTGGGCTTTCGGCTGAGGT 3000  
QY 2963 TCAGGCTTATTTCTAGACTTGCAGTGCAGCTTCCAGAGCTTCCAGAGCTTCCATCAATATA 3022  
DB 3001 TCAGAGCTGTTTCTGAGTTCAGGTTGAGCAGCTTCCAGAGCTTCCAGAGCTTCCATCA 3060  
QY 3023 CAAGATCTTCTGCTTACGCTTACAGTTCATGCTGATGATGATGATGATGATGATGATG 3082  
DB 3061 CAAGATCTTCTGCTTACGCTTACAGTTCATGCTGATGATGATGATGATGATGATGATG 3120  
QY 3083 CAGGCTGTTTGAAGAACCTTACATCTTCTGCGCATATCTCCAGCAGCATCTGCT 3142  
DB 3121 TCAGAGAGTTTGAAGAACCTTACATCTTCTGCGCATATCTCCAGCAGCATCTGCT 3180  
QY 3143 CTGCTATGCTATCTGAGGCTCAAGAACTCCAGGATGACATTAAG-----GCCTC 3193  
DB 3181 CTGCTACTCATCTTGAAGAACCTTACATCTTCTGCGCATATCTCCAGCAGCATCTGCT 3240  
QY 3194 TGCTCTCTTCTCTGAGGCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3283  
DB 3241 CGGCTCTGCTCTGAGGCGGCTGAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3300  
QY 3254 GCTGCTCTTCTCTGAGGCGGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3313  
DB 3301 GACTCGACAGCTGTCAGCTAGTCCACTCTCTGGGGTCACTCAGGACAGCCAGAGCA 3360

QY 3314 GCTGTGCCGGAAGCTCCAGAGCGCAATGACCATCTTAAAGCTGACAGTACCCAGC 3373  
DB 3361 GCTGAGTCGGAAGCTCCGCGGAGCAGCGTACTGCTTGGAGCGCGCAACCCGCG 3420  
QY 3374 CCTAAGCACAGACTTTCAGACCACTTTTGGACTAA 3407  
DB 3421 ACTGCCCTCAGACTTCAAGACCATCTCTGGACTGA 3454

RESULT 2  
V60320  
ID V60320 standard; cDNA; 4023 BP.  
AC V60320;  
DT 04-DEC-1998 (first entry)  
DE human telomerase gene referred to as hEST2.  
KW Catalytic subunit; human; telomerase; telomere maintenance;  
OS diagnosis; treatment; cancer; ss.  
FH Homo sapiens.  
FT Key Location/Qualifiers  
CDS 59..3458  
/\*tag= a

W09837181-A2..  
27-AUG-1998.  
PF 20-FEB-1998; U03404.  
PR 30-OCT-1997; US-064322.  
PR 20-FEB-1997; US-038750.  
PR 20-MAY-1997; US-047151.  
PR 01-AUG-1997; US-054549.  
PR 14-AUG-1997; US-055762.  
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PI Counter CM, Meyerson M, Weinberg RA,  
DR WPI: 98-495367/42.  
DR P-PSDB: W71376.  
PT New isolated human telomerase catalytic sub-unit gene - used to  
develop products for increasing or reducing the life span of cells  
such as cancer cells or transformed cells  
PS Claim 5; Fig 5A-B; 96pp; English.  
CC The present sequence encodes the catalytic subunit of a human  
telomerase holoenzyme. Disruption of the telomerase gene alters  
and the protein is physically associated with telomerase activity,  
constituent of active telomerase complex. The products can be used  
for increasing or reducing the lifespan of cells such as cancer cells  
or transformed cells. They can also be used in the diagnosis and  
treatment of malignancies. In addition, cells with a longer lifespan  
can be transplanted into or grafted onto an individual (e.g. as skin  
grafts, as systems for delivery of therapeutic proteins, such as hormones  
and enzymes), to whom they provide therapeutic benefit.  
SQ Sequence 4023 BP; 668 A; 1363 C; 1277 G; 715 T;

Query Match 45.8%; Score 1600.8; DB 1; Length 4023;  
Best Local Similarity 68.9%; Pred. No. 0;  
Matches 2379; Conservative 0; Mismatches 967; Indels 108; Gaps 9;

QY 23 CCGGGCTTGAAGCAATGACCCGGCTCTCTGTTGCCCGCGGTCTCTCTCTCGG 82  
DB 43 CCGGGCAACCCCGCATGCGCGGCTCCCGCTGCGAGCGCTGCTCTCTCTCGG 102  
QY 83 CAGCCCATACCCGGAGGTGCGCGCTGCGCAACCTTTGTGGGCGCTTGGGCGCCGAGG 142  
DB 103 CAGCCACTACCCGGAGGTGCTCCCGCTGCGCAACCTTTGTGGGCGCTTGGGCGCCGAGG 162  
QY 143 CAGGGGCTTGTGCAACCCGGGACCCGAGATCTACCGCACTTTGTTGCCCAATGCTT 202  
DB 163 CTGGGGCTGTGTGAGCGCGGGACCCGCGGCTTTCGCGCGCTGTTGGGCGCCAGTGCCT 222  
QY 203 AGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 262  
DB 223 GTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 282  
QY 263 ATCCCTGAAGAGCTGTGTGGCCAGGTTGTGACAGACTCTGCGAGCGCAACAGAGAAA 322

Mon Aug 16 10:21:24 1999

Db 283 CTGCTGAAGAGCTGGTGGCCGAGTGTGTCAGAGGCTGTGCGAGCGGGCGGGAAGAA 342  
 QY 323 CGTGTGGCTTTTGGCTTTTGGCTTTTAAAGAGCCAGAGCGGGGCTTCCCATGGCCTT 382  
 Db 343 CGTGTGGCTTTGGCTTTTGGCTTTTAAAGAGCCAGAGCGGGGCTTCCCATGGCCTT 402  
 QY 383 CACTAGTAGGCTGCTAGTACTTGGCCCAACACTGTTATTAAGACCTGCTGCTGCTGCTG 442  
 Db 403 CACACAGCGCTGGCAGCTTACCTGCCAACACAGCTGACCGACGACTGCGGGGAGCGG 462  
 QY 443 TGCATGATGCTACTGTTGAGCGAGTGGCGGACGACTGCTGGTCTACTGCTGGCACA 502  
 Db 463 GGCCTGGGGCTGCTGTTGGCGCGCTGGCGGACGACTGCTGGTCTACTGCTGGCAG 522  
 QY 503 CTGCTGCTTTTATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 562  
 Db 523 CTGCGCGCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 582  
 QY 563 GTACCAATTTGTGCCACACGAGTATCTGGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 622  
 Db 583 GTACAGCTGCGGCTGCGACTCAGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCT 642  
 QY 623 CCGACCTGCGGAGGAATTTCACTAAGCTTAGTCTTACAAAGATCAAGAGCAGTAG 682  
 Db 643 GCGTCTGGGATGC-----GAACGGCGCTTGAACCATAG 675  
 QY 683 TCGCCAGGAAGCAGCAACCCCTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 742  
 Db 676 CGTCAAGGAGCGGGGCTTGGCGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 735  
 QY 743 TCTACCAAGTACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 802  
 Db 736 CAGTCCAGCGGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 795  
 QY 803 GGAGGAGGAGCAGCAGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 862  
 Db 796 GGAGGAGCAGCAGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 855  
 QY 863 TCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 913  
 Db 856 TGACCGTGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 915  
 QY 914 AGGAAAGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 970  
 Db 916 GGTGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 975  
 QY 971 CTCACAT---CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1027  
 Db 976 CCGCCCATCCACATCGCGGCGCCACAGTCCCTGGGACAGCGCTTGTCCCGGCTGCTGCTG 1035  
 QY 1028 TGAGACAGACATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1087  
 Db 1036 CGAGACCAAGCACTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1092  
 QY 1088 CCTACTCAGCAACTCCAGCTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1147  
 Db 1093 CCTACTCAGCTTCTGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1152  
 QY 1148 TCTGGCTCAAGGCTTAGGACTCAGGACCACTCTGCGAGACACACCGTCTATGCGCTG 1207  
 Db 1153 TCTGGTTCAGGCGCTTGGATGCGAGGACTCCCGCAGGTTGCGCGGCTGCGCGGCTGCGCG 1212  
 QY 1208 ATACTGGCAGATGGGCGCTTGTCCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1267  
 Db 1213 CTACTGGCAATGGGCGCTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1272  
 QY 1268 TCTCAGACTCTCAGTGCACATTCAGGTTTCGAACAGCAAAACCAACAGGTGCAGATGC 1327  
 Db 1273 CGGGTGTCTCAGAGCGACTGCCGCTGCGAGCTGCGGTCAACCCAGCAGCGCGGTGT 1332  
 QY 1328 CT-----TGAAACACAGCGCC 1342  
 Db 1333 CTGTGCGCGGAGAGCCCGAGGCTGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1392

QY 1343 ACCGACCTCATGGATTTGCTCCGCTGTCAGCAGAGTCCCTGGCAGGTATATGCTTTCT 1402  
 Db 1393 CGGTGCGCTGGTGCAGTGTCCGCGCAGCAGACGCCCTGGCAGGTGTACGGCTTCGT 1452  
 QY 1403 TCGGGCTCTGCTCTGCAAGTGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1462  
 Db 1453 GCGGGCTGCTGCTGCGCGCTGGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1512  
 QY 1463 CCGCTCTTTTAAAGCTTAAAGAACTTCACTCTGTTGGGAAATACGCAAGCTATCACT 1522  
 Db 1513 CCGCTCTTCCAGCAACACCAAGAGTTTCACTCTCCCTGGGAAAGCATGCAAGCTCTCGCT 1572  
 QY 1523 GCAGGAACTGATGTGGAAGATGAAGTAGAGTGGCACTGCTGCTGCTGCTGCTGCTGCTGCTG 1582  
 Db 1573 GCAGGAGCTGACGTGAAGATGAGCGTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1632  
 QY 1583 GAAGGACGCTGCTCCGCTGCGAGCAGCAGCTGCTGAGGGAGAGATCCTGGCTGCTGCTGCTG 1642  
 Db 1633 GGTGGCTGTGTTCCGCGCGCAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1692  
 QY 1643 GTTCTGGCTGATGAGACACATACGTGGTACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1702  
 Db 1693 GCATGCTGATGATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1752  
 QY 1703 GAGCAGCTCCAGGAAGCAGGCTTCTTCTTACCGTAAAGTGTGCTGCTGCTGCTGCTGCTGCTG 1762  
 Db 1753 GACCAAGTTCCTGCGCGCGCAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1812  
 QY 1763 GAGCATGGAATGAGGAAACACCTTGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1822  
 Db 1813 RAGCATGGAATGAGGAAACACCTTGAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1872  
 QY 1823 GGTAGGATCAGCAGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1882  
 Db 1873 GGTAGGAGATCAGGAGGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1932  
 QY 1883 CAAGCCCAAGCGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1942  
 Db 1933 CAAGCTGAGGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1992  
 QY 1943 GGGCAGGAAGAGCAGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2002  
 Db 1993 CCGCAGAGAAAGAGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2052  
 QY 2003 CAACTATGAGCGGCAAAACATCCTCACTATATGGGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2062  
 Db 2053 CAACTAGAGCGGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2112  
 QY 2063 CATCTACAGACCTGCGCGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2122  
 Db 2113 TATCCACAGGCGCTGCGCGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2172  
 QY 2123 GATGCTACTTTTAAAGCAGATGTGACCGGGGCTTATGATGCTATGCTATGCTATGCTATGCTAT 2182  
 Db 2173 GCTGCTACTTTGCTCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2232  
 QY 2183 GGTGAGGTTTGTGCTCAATATGATCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2242  
 Db 2233 CACGAGGCTCATGCGCAGCATCATCA---AACCCAGAACAGCTGCTGCTGCTGCTGCTGCTG 2289  
 QY 2243 TGCAGTGTCCGAGAGATAGCCAAAGCCAAAGTCCCAAGTCTTATGAGGACAGGTCTAC 2302  
 Db 2290 TCGCGTGTCCAGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2349  
 QY 2303 CACCTCTCTGACCTCCAGCCATACATGGCGGCTTCCCTTAAGCATCTGCAAGGATTCTG 2362  
 Db 2350 TACCTTGACAGACCTCCAGCGCTACATGCGACAGTTCGTGGCTGCTGCTGCTGCTGCTGCTG 2403  
 QY 2363 TCGCAGTGCATGAGGAAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2422  
 Db 2404 GACCAAGCGCTGAGGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2463









Mon Aug 16 10:21:24 1999

QY 503 CTGTGCTCTTTATCTTCTGTCGCCGCCAGCTGTGCTTACAGAGTGTGTGGTCTTCCCT 562  
 DB 532 CTGCGGCTCTTTGTGCTGGTCTCCAGCTGCGCTTACAGAGTGTGTGGTCTTCCCT 591  
 QY 563 GTACCAAAATTTGTGCACACAGGATATCTGGGCTCTGTGCTGCTTACAGGCTTAC 622  
 DB 592 GTACCAAGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 651  
 QY 623 CCAGCCGCTGGCAGGAATTCACCTAAGCTTAGGTTCTTACACAGATCAAGAGCAGTAG 682  
 DB 652 GCCTGTGGGATG-----GAGCGGCTGGAACCATAG 684  
 QY 683 TCAGCAGGAGACCGAAACCCCTGGCTTGGCTTGCATCTGAGGTACAAAGAGCATCTGAG 742  
 DB 685 CGTCAGGAGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 744  
 QY 743 TCTCAGCTACAGTGTGCTTCACTAAGAGGCGAGATGCTATCTTCTGCTGCTGCTGCT 802  
 DB 745 CAGTGCAGCGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 804  
 QY 803 GGAGGAGGAGACCCACAGGAGTGTACCAACCCCATAGGCAATCATGGTGGCAGAG 862  
 DB 805 GGAGCGAGCGGCTTGGCAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 864  
 QY 863 TCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 913  
 DB 865 TGACCGTGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 924  
 QY 914 AGGAAAGTGTCTGACCTGAGTCTCTC-----GCAGAGAAAGATTTGCTTCTTAA 970  
 DB 925 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 984  
 QY 971 CTCCACAT---CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1027  
 DB 985 CCCCCATCCACATCGCGGCGACACAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1044  
 QY 1028 TCAGACAGACATTTCTTTACTCCAGGAGATGGCCAGAGGCTCTAAACCCCTCAT 1087  
 DB 1045 CAGACACAGACATTTCTTCTTACTCTCTCAGGCA---CAAGGAGCAGCTGCGGCTCT 1101  
 QY 1088 CTTACTAGCAACCTCCAGCTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1147  
 DB 1102 CTTACTGCTCTCTGAGGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1161  
 QY 1148 TCTGGCTCAAGCTAGGATCAGGACCTCTGAGGACACACCTCTATCGCTG 1207  
 DB 1162 TCTGGCTTCCAGGCTGATGCTGAGGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1221  
 QY 1208 ATACTGGCAGATCGGCGCTGTTCCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1267  
 DB 1222 CTACTGGCAATGCGGCTGTTCTTGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1281  
 QY 1268 TGTCACTCTCAGTCACTGAGTTCGAGGTTTCGAACAGCAACCAACAGGTGACAGTGC 1327  
 DB 1282 CGGGTGTCTTCAACAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1341  
 QY 1328 CT-----TGACACAGGCT 1342  
 DB 1342 CTGTGCGCGGAGAACCCAGGCTCTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1401  
 QY 1343 ACCGACCTCATGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1402  
 DB 1402 CCGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1461  
 QY 1403 TCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1462  
 DB 1462 CGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1521  
 QY 1463 CCGCTTCTTAAAGACTTAAAGACTTAAAGACTTAAAGACTTAAAGACTTAAAGACTTAA 1522  
 DB 1522 CCGCTTCTCAGGAACACCAAGAGTTCATCTCCCTGCGGAGCATGCCAGCTCTGCT 1581  
 QY 1523 GCAGGAATGATGTGGAAGATGAAAGTAGAGATTGCCACTGCTGCTGCTGCTGCTGCTG 1582

DB 1582 GCAGGAGCTGAGTGAACATGAGCGTGGGACTGCGCTTGGCTGCGCAGAGGCCAGG 1641  
 QY 1583 GAAGGACCGTGTCCCGCTGCGAGACCGCTGTGAGGAGAGGATCTTGGCTTACCTTCT 1642  
 DB 1642 GGTGGCTGTGTCCCGCGCAGAGCAGCTGTGCTGAGGAGATCTTGGCCAACTTCT 1701  
 QY 1643 GTTCTGGCTGATGGACACATACGTGTACAGCTGTGTAGTTCATTTTATACATCAGA 1702  
 DB 1702 GCATGGCTGATGAGTGTGTACGTGCTGAGTGTGTACGTCTTCTTTATGTCAAGA 1761  
 QY 1703 GAGCACATCCAGAGAACAGGCTCTTCTTACCTTAAGAGTGTGTGAGCAAGCTGCA 1762  
 DB 1762 GACCACGTTTCAAAAGAACAGGCTCTTCTTACCGGAGAGAGTGTGTGAGCAAGTGA 1821  
 QY 1763 GAGCATTTGAGTCAAGCAACACCTTGAAGAGTGTGAGTGTGAGTGTGAGTGTGAG 1822  
 DB 1822 AAGCATTTGAGTCAAGCAACACCTTGAAGAGTGTGAGTGTGAGTGTGAGTGTGAG 1881  
 QY 1823 GGTGAGGCTACACAGACACCTGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1882  
 DB 1882 GGTGAGGCTACACAGACACCTGCTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTG 1941  
 QY 1883 CAAGCCCAAGGCTGCGGCGCTTGTGAACATGAGTGTATAGCATGGGTACAGAGCTTT 1942  
 DB 1942 CAAGCTTACAGGCTGCGGCGCTTGTGAACATGAGTGTATAGCATGGGTACAGAGCT 2001  
 QY 1943 GGGCAGAGAGGAGGCGGCGCTTGTGAACATGAGTGTATAGCATGGGTACAGAGCTTT 2002  
 DB 2002 CGCGAGAGAGAGGCGGCGCTTGTGAACATGAGTGTATAGCATGGGTACAGAGCT 2061  
 QY 2003 CAATCTAGCGGAGAGAGGCGGCGCTTGTGAACATGAGTGTATAGCATGGGTACAG 2062  
 DB 2062 CAATCTAGCGGAGAGGCGGCGCTTGTGAACATGAGTGTATAGCATGGGTACAGAG 2121  
 QY 2063 CATCTACAGGAGGCGGCGCTTGTGAACATGAGTGTATAGCATGGGTACAGAGCTTT 2122  
 DB 2122 TATCCACAGGCGGCGCTTGTGAACATGAGTGTATAGCATGGGTACAGAGCTTT 2181  
 QY 2123 GATGTACTTGTAGGAGAGGCGGCGCTTGTGAACATGAGTGTATAGCATGGGTACAG 2182  
 DB 2182 GCTGTACTTGTAGGAGAGGCGGCGCTTGTGAACATGAGTGTATAGCATGGGTACAG 2241  
 QY 2183 GGTGAGGCTTGTGCAATATATGATCAGGCTTGTGAGAGGAGGCTTGTATCCGCG 2242  
 DB 2242 CAGGAGGCTATCGGCGGCTTGTGAGAGGAGGCTTGTATCCGCGCTTGTATCCG 2298  
 QY 2243 TGAGTGTGCGGAGAGATAGCCAGGCGCTTGTGAGAGGAGGCTTGTATCCGCGCTTGT 2302  
 DB 2299 TGCGGTGCGGAGAGGCGGCGCTTGTGAGAGGAGGCTTGTATCCGCGCTTGTATCCG 2358  
 QY 2303 CACCTCTCTGAGCTCCAGGCTTGTGAGAGGAGGCTTGTATCCGCGCTTGTATCCG 2362  
 DB 2359 TACCTTGAAGAGCTCCAGGCTTGTGAGAGGAGGCTTGTATCCGCGCTTGTATCCG 2412  
 QY 2363 TGCGAGTGTACTGAGGAGGCTTGTGAGAGGAGGCTTGTATCCGCGCTTGTATCCG 2422  
 DB 2413 GAGCAGGCGCTGAGGAGGCTTGTGAGAGGAGGCTTGTATCCGCGCTTGTATCCG 2472  
 QY 2423 CAGCAGGCTTGTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTGAG 2482  
 DB 2473 CAGTGGCTTGTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTGAG 2532  
 QY 2483 CAGTGTCTATGAGGAGGCTTGTGAGAGGAGGCTTGTGAGAGGAGGCTTGTGAGAG 2542  
 DB 2533 CAGTGTCTATGAGGAGGCTTGTGAGAGGAGGCTTGTGAGAGGAGGCTTGTGAGAG 2592  
 QY 2543 CAGTGTCTATGAGGAGGCTTGTGAGAGGAGGCTTGTGAGAGGAGGCTTGTGAGAG 2602  
 DB 2593 CAGTGTCTATGAGGAGGCTTGTGAGAGGAGGCTTGTGAGAGGAGGCTTGTGAGAG 2652  
 QY 2603 GCTTTAGCTTTTGTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTGAGTGTCTG 2662



1510 CGCCTTCTCAGGAACACCAAGAAAGTTCTCTCCCTGGGAAGCATGCCAAGCTCTCGCT 1569  
 1523 GCAGGAACTGATGTGAAGATGAAGATAGAGATTGCCACTGGCTCCCGCAGACGCCGGG 1582  
 1570 GCAGGAGCTGACGTGGAAGATGAGCGTGGGAGTGGCTTGGCTGGCGAGGAGCCAGG 1629  
 1583 GAAGGACCGTGTCCCGCTCCAGAGACCGCTGAGGAGAGATCCCTGGCTACGTTCTCT 1642  
 1630 GGTGGCTGTGTCCCGCGCAGAGACCGCTCTGCTGAGAGATCCCTGGCAAGTCTCT 1689  
 1643 GTTCTGGCTGATGAGACACATACGCTGATGAGAGTGTAGTCTATTTTATACATCAGA 1702  
 1690 GCATGGCTGATGAGTGTGATGCTGCTGAGCTGCTGAGTCTTTTATGTCACGGA 1749  
 1703 GAGCACATTTCCAGAAACAGAGCTCTTTCTTACCGTAAAGAGTGTGGAGCAAGTGA 1762  
 1750 GACCAGCTTTCAAAGAAACAGGCTCTTTTCTACCGCGCGAGTGTCTGGAGCAAGTGA 1809  
 1763 GAGCATTTGGAGTCAAGCAACACCTTGAGAGATGCGGCTACGGAGCTGTCAACAAGGA 1822  
 1810 AAGCATTTGAATCAGACAGCACCTTTGAAGAGGGTGCAGCTCGGAGCTGTGGAAGCAGA 1869  
 1823 GGTGAGGATCACCAGGACACCTTGGCTAGCCTAGCCTATCTGCAGACTGCGCTTATFCCC 1882  
 1870 GGTGAGGATCAGGAGCAACCTTGAGAGATGCGGCTACGGAGCTGTGGAAGCAGA 1929  
 1883 CAAGCCCAACCGGCTGCGGCGCTTGTGAACATGATGATGATGATGATGATGATGATGAT 1942  
 1930 CAAGCTGACGCGGCTGCGGCGCTTGTGAACATGATGATGATGATGATGATGATGATGAT 1989  
 1943 GGGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2002  
 1990 CCGCAGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2049  
 2003 CAACTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2062  
 2050 CAACTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2109  
 2063 CATCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2122  
 2110 TATCCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2169  
 2123 GATGTACTTTGTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2182  
 2170 GCTGTACTTTGTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2229  
 2183 GGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2242  
 2230 CACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2286  
 2243 TGCAGTGGTCCCGAGAGATAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2302  
 2287 TGCCTGGTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2346  
 2303 CACCTCTCTGAGCTCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2362  
 2347 TACCTTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2400  
 2363 TGCAGTGGTCCCGAGAGATAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2422  
 2401 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2460  
 2423 CAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2482  
 2461 CAGTGGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAG 2520  
 2483 CAGTGGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAG 2542  
 2521 CAGTGGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAG 2580  
 2543 CAGTGGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAGCTCTTCCAG 2602  
 2581 CAGCTGTCTGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2640

Qy	2603	GCTTTTACGTTTTTGTGATGACATTTCTGTTGGTGAGCCTCACTTTGGACCAAGCAAAAC	2662
Db	2641	GCTCCTCGGTTTGGTGATGATTTCTGTTGGTGACACCTCACTTACCACGGGAAAC	2700
Qy	2663	CTTCTCAGACACCTGCTCATGGCGTTCCTGAGTATGGTGCATGATAAACTTTCGAA	2722
Db	2701	CTTCTCAGGACACCTGCTCGAGGCTGCTCCTGAGTATGGTCTGGTGGTGAACCTTTCGAA	2760
Qy	2723	GACAGTGGTGAACATTCCTGTGGAGCCTGGTACCTGGTGGTGGAGCTCCATACCAAGCT	2782
Db	2761	GACAGTGGTGAATTCCTGTGAAGACAGAGCCCTGGTGGCAGCGCTTTGTTCAGAT	2820
Qy	2783	GCCTGCTACTGCGCTGTTTCCCTGGTGTGGCTTGTGTGGACACTCAGACTTTGGAGGT	2842
Db	2821	GCCGGCCACGGGCTATTCCCTTGGTGGCGCCTGCTGTGGATACCCGGACCTGGAGGT	2880
Qy	2843	GTTCTGTGACTACTCAGGTTATGCCAGACACTCAATTAAGACGAGCCTCACTTCCAGAG	2902
Db	2881	GCAGAGCGACTACTCAGCTATGCCCAGGACCTCCATCAGAGCCAGTGTCACTTCAACCG	2940
Qy	2903	TGCTTCCAAAGCTGGGAAGACCATCGGAACAAGCTCCTGCTCGCTTGTGGGTGAAGTG	2962
Db	2941	CGGCTTCAAGGCTGGGAGGAACATCGTCGCAAACTCTTTGGGCTCTGGCGTGAAGTG	3000
Qy	2963	TCAGGCTATTCTAGACTTCGAGTGAACAGCCTCCAGACAGTCTGCATCAATATATA	3022
Db	3001	TCAGAGCCTGTTCTGGATTTTCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTA	3060
Qy	3023	CAAGATCTTCTCCTTCAGGCGCTACAGGTTCCATGCATGTGTGATTTCAGCTTCCCTTTGA	3082
Db	3061	CAAGATCCTCCTGCTCAGGCTACAGGTTTCACGCATGTGTGTGCGAGCTCCCATTTCA	3120
Qy	3083	CCAGCGTGTTAGGAAGACCTCACATTTCTTGGCATCATCTCCAGCCAAAGCATCCTTG	3142
Db	3121	TCAGCAAGTTTGGGAAGACCCACATTTTCTCGGGGTCACTCTGCACAGGCCCTCCCT	3180
Qy	3143	CTGCTATGCTATCCTGAAGGTCAAGAATCCAGGAATGACACTAAG-----GCCTC	3193
Db	3181	CTGCTACTCCATCTGAAAGCCAAAGCGAGGATGTGCTGGGGCCAAAGGGCGCCGC	3240
Qy	3194	TGGCTCCTTTCCTCCTGAAGCGGCACATTTGGCTCTGTGTACAGGCGCTTCTGCTCAAGCT	3253
Db	3241	CGGCGCTCTGCGCTCCGAGGCGGTGCACTGGCTGTGCCACCAAGCATTCCTGCTCAAGCT	3300
Qy	3254	GGCTGCTCATTTCTGTCACTACAATGTCCTCTGGGACCTCTGAGGACGCCCAAAACT	3313
Db	3301	GACTGCACACCGTGTCACTACGTGCCACTCTGGGGTCACTCAGGACAGCCCGCAGCGCA	3360
Qy	3314	GCTGTGCCGGAAGCTCCCGAGAGGGACAAATGACCATCCTTTAAAGCTGCAGCTGACCCAGC	3373
Db	3361	GCTGAGTCGGAAGCTCCCGGGGAGCAGCTGACTCTGGGGTCACTGGAGGCCGACCAACCCGC	3420
Qy	3374	CCTAAGCACAGACTTTCAGACCATTTTGGACTAA	3407
Db	3421	ACTGCCCTCAGACTTTCAGAGACCATCTCTGGACTGA	3454

RESULT 6

ALCOU  
V22382

ID V22382 standard; cDNA: 3855 bp.

AC V22382;

NT 13-AUG-1998 (first entry)

Human telomerase reverse transcriptase Delta182 variant encoding cDNA

RW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;  
prognosis; cell proliferation

RW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;

RW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;  
prognosis; cell proliferation

prognosis; cell proliferation; cancer; ageing; ribonucleoprotein; ss. synthesis

OS  
synthetic.  
OS Homo sapiens.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT CD\$  
56. .2479  
Production/Quartiers

```
FT tag= a
```

```

FT      /product= "telomerase reverse transcriptase variant"
DN      CCCCCTTCTGGCT

```

PN GB2317891-A.

[illegible]

Q B M

Query Match	Score	DB 1	Length
37.88	1323.2	DB 1	Length 3855

Best Local Similarity 65.2%; Pred. No. 0;  
Matches 2251; Concentration 0.000000

QY Db QY Db QY Db QY Db QY Db QY Db QY









[illegible][illegible]





```
Db 892 GACACCTACCTCACCCAGCGAAACCTTCTCAGGACCCTGCTGAGGTGTCCTCTGA 951
QY 2696 GTATGGTGCATGATAAATTTGCGAGAGACAGTGGTGAACCTTCCCTGTGGAGCCTGTGAC 2755
Db 952 GTATGGTGCCTGGTGGTGAATTTGCGAAGACAGTGGTGAATTCCTGTAGAGAGAGGC 1011
QY 2736 CTTGGTGGTGCAGCTCCATCACCAGCTGCTGCTCACTGCTGCTTTCCCTGGTGTGCTT 2815
Db 1012 CTTGGTGGTGCAGCTTTTGTTCAGATGCCGCGCCACGGGCTATTCCTCCTGGTGGCCT 1071
QY 2816 GCTGCTGCACACTCAGACTTTGGAGGTGTTCTGTGACTACTCAGGTTATGCCAGACCTC 2875
Db 1072 GCTGCTGGATACCGGACCTTGGAGGTGCAGAGCGACTACTCCAGCTATGCCCGGACCTC 1131
QY 2876 AATTAAGACGAGCTCACCCTCCAGAGTGTCTTCAAGCTGGGAAGACCATGCGGAACAA 2935
Db 1132 CATCAGAGCGAGTCTACCTTCAACCGGCTTCAAGGCTGGGAGGAACATGCTGCGCAA 1191
QY 2936 GCTCCTGTGGTCTGTCGCTTGAAGTGTACAGGTCTATTTCTAGACTTGCAGGTGAACAG 2995
Db 1192 ACTCTTTGGGCTCTTGGGCTGAAGTGTACAGCTGTTCTGAGTTTGCAGGTGAACAG 1251
QY 2996 CTTCCAGACAGTCTGCATCATATATACAGATCTTCTGCTTACGGCCTACAGGTTCCTCA 3055
Db 1252 CTTCCAGACAGTGTGCACCAACATCTACAGATCTTCTGCTGAGGCTGACAGTTTCA 1311
QY 3056 TGCATGTGTGATCAGCTTCCCTTTGACGAGCTGTAGAGAGACCTCACATTTCTTCT 3115
Db 1312 CGCATGGTGTGCAGCTCCCATTTTCATCAGCAAGTTTGGAGAACCCACATTTTCTCT 1371
QY 3116 GGGCATCATCTCCAGCAAGCATCTGCTGCTATGCTATCTTCAAGGTCAAGAAATCCAGG 3175
Db 1372 GCGGTCATCTGTGACAGCGGCTCCCTGCTACTCTCATCTCCAGCAAGCAAGCAAGG 1431
QY 3176 AATG 3179
Db 1432 ATG 1435

RESULT 10
X15925
ID X15925 standard; cDNA; 2175 BP.
AC X15925;
DT 14-MAY-1999 (first entry)
DE Nucleotide sequence of the specification.
KW Catalytic telomerase; diagnosis; disease; telomerase activity; ds.
OS Homo sapiens.
PN J11046768-A.
PD 23-FEB-1999.
PF 01-AUG-1997; 207708.
PR (MITU ) MITSUBISHI CHEM CORP.
DR WPI: 99-208111/18.
PT New catalytic protein of telomerase of a higher animal and a gene
PT coding it - useful for diagnosis of diseases caused by the change in
PT activity of a telomerase
PS Example 2; Page 15; 18pp; Japanese.
CC The specification describes a human catalytic telomerase protein.
CC The products are useful in drug compositions for the diagnosis
CC of diseases caused by the change in activity of telomerase. The
CC present sequence appears in the specification.
SQ Sequence 2175 BP; 433 A; 679 C; 641 G; 422 T;

Query Match 20.5%; Score 715; DB 1; Length 2175;
Best Local Similarity 66.4%; Pred. No. 1.7e-195;
Matches 1185; Conservative 0; Mismatches 400; Indels 200; Gaps 3;

QY 1632 GCTACGTTCTGTTGGTGTGATGAGACATACGTGCTGACGCTTATAGTCAATCTTT 1691
Db 1 GCCAAGTCTCTGCACTGGTGTGAGTGTGACGTGCTGACGTGCTCAGGCTTCTTT 60
QY 1692 TACATCACAGAGACATCTCCAGAGAACAGGCTCTTCTTCTACCGTAAGAGTGTGG 1751
```

```
Db 61 TATGTCAGGAGACACGCTTTCAAAGAACAGGCTCTTTTCTACCGAAGAGTGTCTGG 120
QY 1752 AGCAAGCTGCGAGCATTTGGAGTCAAGCAACACCTTGAGAGAGTGGGCTACGGAGCTG 1811
Db 121 AGCAAGTTGCAAGCATTTGGAATCAGACGACCTTGAGAGAGGTTGACGTCGGGAGCTG 180
QY 1812 TCACAAGAGGAGGTCAAGCATCACCAGGACACCTGGGTAGCCATGCCATCTGCAGACTG 1871
Db 181 TCGGAAGCAGAGGTCAAGCAGCATCGGGAAGCCAGGCCCTGCTGAGCTGCAGACTC 240
QY 1872 CGCTTCATCCCAAGCCCAAGCGCTGGGCGCATTTGTGAACATAGTTATAGCATGGGT 1931
Db 241 CGCTTCATCCCAAGCGCTGAGCGGCTCGGCGCATTTGTGAACATAGTTATAGCATGGGT 1931
QY 1932 ACCAGAGCTTTGGCAGGAAGAGCAGGCGCAGCATTTACCCAGCGCTCTCAAGACTCTC 1991
Db 301 GCCAGAAGCTTCCCGCAGAGAAAGAGGCGCGAGGCTCTACCTCGAGGGTGAAGGCACTG 360
QY 1992 TTCAGCATGCTCAACTATGAGCGGACAAAACATCTCACCTTATGGGGTCTTCTGTACTG 2051
Db 361 TTCAGCTGTTCAACTACGAGCGGCGCGCGCGCTCTCTGGCGCTCTGTGCTG 420
QY 2052 GGTATGAATGACATCTACAGGACCTGCGGCGCTTTTGTGCTGGGTGCTGCTGCTGCTG 2111
Db 421 GGCCTGGACCATATCCACAGGCGCTGGCGACCTTCTGCTGCTGCGGCGCGGAGGAC 480
QY 2112 CAGACACCCAGAGTACTTTTGAAGGAGATGTACCGGGGCGCTTATGATGCCATCCCC 2171
Db 481 CCGCGGCTGAGTGTACTTTGCAAGTGGATGTGACGGGCGCTGACGACCATCCCC 540
QY 2172 CAGGGTAAAGTGTGAGGTTGTGCAATATGATCAGGCACTCGGAGAGCAGTACTGT 2231
Db 541 CAGGACAGCTCAGGAGGTCTCGCCAGCATCATCA---AACCACAGAACACGTACTGC 597
QY 2232 ATCCGCGCATGATGAGTGTGCGGAGAGATAGCCAAAGGCGCAAGTCCACAAGTCTTAGG 2291
Db 598 GTGCTGCGGTATGCGGTGCTCCAGAAGGCGCGCATGGCAGCTCCGCAAGGCTTCAAG 657
QY 2292 AGACAGGTCACACCCCTCTCTGACCTCCAGCCATACATGGGCCAGTTCCTTAAAGCATCTG 2351
Db 658 AGCCAGCTC----- 667
QY 2352 CAGGATTCAGATGCCAGTGCACGTAGAGAACTCCGTTGTCATCGAGCAGAGCATCTCTATG 2411
Db 667 ----- 667
QY 2412 AATGAGCAGCAGCAGCTGTTTGTACTTCTCTGCACTTCTCTGCTCACAGTGTGTA 2471
Db 667 ----- 667
QY 2472 AAGATTGTTGACAGTGTCTATACGAGTCCAGGCGCATCCCCAGGCGCTCCAGCTATCC 2531
Db 667 -----CTAGTCCAGTGCAGGGGATCCCGCAGGCGCTCATCTCTCC 709
QY 2532 ACCCTGCTCTGAGTGTGTTTCGAGACATGGAGAACAGCTGTTGCTGAGGTGTCAG 2591
Db 710 ACGTGTCTCTGACGCTGTGTCAGCGGACATGGAGAACAGCTGTTGCGGGGATTCGG 769
QY 2592 CGGGATGGTGTGCTTTTACGTTTGTGATGACTTCTGTTGTGAGCGCTCACTTTGGAC 2651
Db 770 CGGACGGGCTGCTCTGCGTTTGGTGGATGATTTCTTGTGTGACACCTCACCTCACC 829
QY 2652 CAAGCAAAAACCTTCTCAGACACCTGGTCCATGGGCTTCTCTGAGTATGGGTGCATGATA 2711
Db 830 CACGCAAAAACCTTCTCAGGACCTTGGTCCGAGGTGTCCTGAGTATGGCTGCGGTG 889
QY 2712 AACTTGCAGAGACACTGGTGAACCTTCCCTGTGGAGCCCTGGTACCCCTGGGTGTCAGCT 2771
Db 890 AACTTGCAGAGACAGTGGTGAACCTTCCCTGTAGAGAGAGGAGCCCTGGGTGTCAGCT 949
QY 2772 CCATACAGCTGCTGCTCACTGCTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2831
```

Pure and recombinant human Telomerase Reverse Transcriptase and its variants - are useful in the diagnosis, prognosis and treatment of cell proliferation conditions especially cancer and ageing

PS Example 1; Fig 18; 387pp; English.  
 CC The present sequence represents a human telomerase reverse transcriptase  
 CC (hTERT) cDNA clone from the present invention. The present invention also  
 CC describes the following methods: (A) determining whether a test compound  
 CC is a modulator of hTERT, by detecting the change in hTERT recombinant  
 CC protein or polynucleotide, on administration of the compound; (B)  
 CC preparation of recombinant telomerase by contacting a protein  
 CC preparation of hTERT with a telomerase RNA component; (C) detection of  
 CC the hTERT RNA or protein in a sample by binding a relevant probe to the  
 CC sample and detecting the complex formed or in the case of RNA detection,  
 CC amplifying the product and correlating the presence of complex or  
 CC amplification product with presence of hTERT in the sample; and (D)  
 CC increasing the proliferation of a vertebrate cell by increasing hTERT  
 CC expression; and (E) the use of an agent that causes an increase in cell  
 CC ageing. A protein preparation to create a medicament that inhibits  
 CC vertebrate cell proliferation to create a medicament for inhibiting the  
 CC hTERT can be used in the manufacture of medicaments for inhibiting the  
 CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
 CC used to treat conditions that are associated with high telomerase  
 CC activity. A protein preparation of hTERT can also be used in the new  
 CC methods.

methods. 2176 BP. 432 A: 679 C; 641 G; 422 T; sequence

Query Match 20.4%; Score 714.4; DB 1; Length 2176;  
Best Local Similarity 66.3%; Pred. No. 2.6e-195;  
Matches 1185; Conservative 0; Mismatches 401; Indels 200; Gaps 3;

[illegible]

Db.	950	TTTGTTCAGATCCCGGCCACCGGCTATTCCCTGTGTGGGGCTGTCTGTGGATACCCGG	1009
QY	2832	ACTTTGGAGGTGTTCTGTGACTACTCAGTTATGCCACGACCTCAATTAAAGACGACCTC	2891
Db	1010	ACCTTGGAGTGCAGAGGCACTACTCAGCTATGCCCCGGACCTCCATCAGACCACTCTC	1069
QY	2892	ACCTTCCAGAGTGTCTTCAAAGCTGGGAAGACCATCGGGAACAGCTCCTGTGCTCTTG	2951
Db	1070	ACCTTCAACCGCGGCTTCAAGCTGGGAGGAACATGCTGCGCAACTCTTTGGGGTCTTG	1129
QY	2952	CGTTTGAAGTGTCAAGGCTATTTCTAGACTTGCAGTGAACAGCTCCAGACAGTCTGC	3011
Db	1130	CGGCTGAAGTGCACAGCCTGTTTCTGATTTGCAGTGAACAGCTCCAGACGCTGTGC	1189
QY	3012	ATCAATATATACAAAGATCTTCTGCTTCCAGGCTTACAGGTTCCATGTCATGTGTGATTCA	3071
Db	1190	ACCAACATCTACAAGATCTCTCTGCTGCGAGGGGTACAGGTTTCAAGCATGTGTGCTGCAG	1249
QY	3072	CTTCCCTTTTACCACAGCGTGTTAGGAAGAACCTTCAATCTTTCTTGGGCACTATCTCCAGC	3131
Db	1250	CTCCCATTTTCATCAGCAAGTTTGGGAAGAACCCACATTTTCTTGCCTGCATCTCTGAC	1309
QY	3132	CAAGCATCTCTGTGCTATGCTATCTCTGAAGTCAAGATCCAGAAATGACACTAAAG---	3189
Db	1310	ACGGCCTCCCTCTGCTACTCCATCTCTGAAGCAAGAACGAGGAGATGTCCGTGGGGCC	1369
QY	3189	-----GCCCTGGCTCTTCTCTCTGAAGCGCACATTTGGCTCTGCTACCAAGGCTTC	3242
Db	1370	AAGGGCGCGCGCGGCCCTCTGCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTC	1429
QY	3243	CTGCTCAAGTGGCTGCTCATCTGTCTATCAAAATGTCTTCCCTGGGACCTCTGAGGACA	3302
Db	1430	CTGCTCAAGTGTACTGCACACCGTGTCACTACGTGGCCACTCTGGGGTCACTCAGGACA	1489
QY	3303	GCCCAAAACGTGTGCGCGGAGCTCCAGAGGGGACAAATGACCATCCTTTAAAGCTGCCA	3362
Db	1490	GCCCAAGCGCASCCTGAGTGGGAAGTCCCGGGGACGACGCTGACTGCCCTGGAGCGGCA	1549
QY	3363	GCTGACCCAGCGCCTAAGCACAGACTTTTCAGACCATTTTGGACTAA	3407

RESULT 11

V22380

ID	V22380 standard; CDNA, 2176 bp
AC	V22380;
DT	13-AUG-1998 (first entry)

Human telomerase reverse transcriptase cDNA clone /12302.  
Human; telomerase reverse transcriptase; hTERT; TERT; diagnosis;  
prognosis; cell proliferation; cancer; ageing; ribonucleoprotein; ss.  
Homo sapiens.

Key	Location/Qualifiers
1)S	HOMO sapiens.

EH	Key	CDs
ET		23.802

```

E1      ...
E2      ...
E3      ...
E4      ...
E5      ...
E6      ...
E7      ...
E8      ...
E9      ...
E10     ...
E11     ...
E12     ...
E13     ...
E14     ...
E15     ...
E16     ...
E17     ...
E18     ...
E19     ...
E20     ...
E21     ...
E22     ...
E23     ...
E24     ...
E25     ...
E26     ...
E27     ...
E28     ...
E29     ...
E30     ...
E31     ...
E32     ...
E33     ...
E34     ...
E35     ...
E36     ...
E37     ...
E38     ...
E39     ...
E40     ...
E41     ...
E42     ...
E43     ...
E44     ...
E45     ...
E46     ...
E47     ...
E48     ...
E49     ...
E50     ...
E51     ...
E52     ...
E53     ...
E54     ...
E55     ...
E56     ...
E57     ...
E58     ...
E59     ...
E60     ...
E61     ...
E62     ...
E63     ...
E64     ...
E65     ...
E66     ...
E67     ...
E68     ...
E69     ...
E70     ...
E71     ...
E72     ...
E73     ...
E74     ...
E75     ...
E76     ...
E77     ...
E78     ...
E79     ...
E80     ...
E81     ...
E82     ...
E83     ...
E84     ...
E85     ...
E86     ...
E87     ...
E88     ...
E89     ...
E90     ...
E91     ...
E92     ...
E93     ...
E94     ...
E95     ...
E96     ...
E97     ...
E98     ...
E99     ...
E100    ...

```

FT /product= "telomerase reverse transcriptase"  
FT /product= "telomerase reverse transcriptase"

PN GB2317891-A.

PD 08-APR-1998.

01-OCT-1997; 020890.

PR 14-AUG-1997; US-915503.

PR 01-OCT-1996; US-724643.

PR 18-APR-1997; US-844419.

PR 25-APR-1997; US-846017.

PR 06-MAY-1997; US-851843.

PR 09-MAY-1997; US-854050.

PR 14-AUG-1997; US-911312.  
14-AUG-1997; HC-913051

PR 14-AUG-1997; US-912931.  
CERON CORP

PA (GERU-) GERUN CORP.  
UNIV TECHNOLOGY

PA (UYTE=) UNIV TECHNOLOGI CORE  
Andrews WH Cech TR Chapman

PI Andrews WH, Cecil IR, Chapman AS, Harley CB;  
PI Morin CB Nakamura T, Harley CB;

PI MOLT II GB, NAKAMURA T, NAKAMOTO T,  
DB WPT: 98-171633/16.

DR WPJ; 98-171033/10.  
DR P-PSDB: W46998.

UK F-PSDB, W40550:









CC associated with telomerase and a constituent of active telomerase  
CC complex. The products can be used for increasing or reducing the  
CC lifespan of cells such as cancer cells or transformed cells. They  
CC can also be used in the diagnosis and treatment of malignancies.  
CC In addition, cells with a longer lifespan can be transplanted into  
CC or grafted onto an individual (e.g. as skin grafts, as systems for  
CC delivery of therapeutic proteins, such as hormones and enzymes), to  
CC whom they provide therapeutic benefit.  
SQ Sequence 3346 BP; 617 A; 1027 C; 965 G; 735 T;

Query Match 18.5%; Score 647; DB 1: Length 3346;  
Best Local Similarity 66.7%; Pred. No. 7.7e-176;  
Matches 1062; Conservative 0; Mismatches 340; Indels 191; Gaps 2;

QY 1582 GGAGGACCGTGTCCCGCTGACAGACACCGTCTGAGGGAGAGATCCTGGCTACGTTCC 1641  
DB 971 GGGTTGGCTGTGTCCGGCCGACAGACCGTCTGCTGAGGAGATCCTGGCCAAAGTTCC 1030  
QY 1642 TGTCTGGCTGATGGACACATAGTGTGACAGCTGCTTAGTGTATCTTTTACATCACAG 1701  
DB 1031 TGCACTGGCTGATGAGTGTAGCTGTGAGCTGTCTAGGTCTTTCTTTTATGTACCG 1090  
QY 1702 AGAGCATATCCAGAGAACAGCGCTCTTCTTACCGTAAGAGTGTGTGAGCAAGCTGC 1761  
DB 1091 AGACCAGGTTTCAAAGAACAGCGCTCTTTCTACCGAAGAGTGTCTGGGCAAGTTGC 1150  
QY 1762 AGAGCATGTGAGTACGGAACACCTGTGAGAGTGGGGTACGGAGCTGTCAACAGAG 1821  
DB 1151 AAAGCATTTGGAATCAGACAGCATTTGAAGAGGTGCAGCTGCGGAGCTGTGCGAAGCAG 1210  
QY 1822 AGGTGAGGATACACAGGACCTGTGCTAGCCATGCCATCTGCACACTGCGTTTCATCC 1881  
DB 1211 AGGTGAGGACATCGGGAAGCCAGCGCCCGCTGTGACGTCCAGACTCGGCTTCATCC 1270  
QY 1882 CCAAGCCCAAGCGCTCGCGCCCATTTGTGAACATAGTTATAGCATGGTGTACCAAGCTT 1941  
DB 1271 CCAAGCCTGACGGCTGTGCGCGATTGTGAACATGACTACGTGTGGAGCCAGAACGT 1330  
QY 1942 TGGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2001  
DB 1331 TCGGAGAGAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1390  
QY 2002 TCAACTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2061  
DB 1391 TCAACTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1450  
QY 2062 ACATCTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2121  
DB 1451 ATATCCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1510  
QY 2122 GGATGTACTTTGTTAAGGAGATGTGACCGGGCCCTATGATGCCATCCCCAGGTAAGC 2181  
DB 1511 AGCTGTACTTTGTTAAGGAGATGTGACCGGGCCCTATGATGCCATCCCCAGGTAAGC 1570  
QY 2182 TGGTGGAGGTTGTTCCATATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2241  
DB 1571 TCACGAGGATCATCGCCAGCATCATCA---AACCCAGAGACAGTACTGCGTGGCTCGGT 1627  
QY 2242 ATGCAAGTGTCCGGAGAGATACCAAGGCAAGTCCCAAGTCTTTTAGGAGACAGGTCA 2301  
DB 1628 ATGCGGTGTCCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1687  
QY 2302 CCACCCCTCTCTGACCTCCAGCATATACATGGGCGCATCTTCTTAAGCATCTGAGGATTCAG 2361  
DB 1687 ----- 1687  
QY 2362 ATGCCAGTGCATGAGGAACCTCCGTTGTATCCGAGAGCATCTCTATGAATGAGACCA 2421  
DB 1687 ----- 1687  
QY 2422 GCAGCAGCGCTGTTGACTTCTCTGCACTTCCCTGCACTGCTGAGGAGGAGGAGGAGGAG 2481

DB 1687 ----- 1687  
QY 2482 ACAGGTGCTATACGAGTCCAGGAGCATCCCCAGGGTCCAGGCTTATCCACCCGTCTCT 2541  
DB 1687 -----CTACGTCCAGTCCAGGAGGATCCCCAGGGTCCAGGCTTATCCACCCGTCTCT 1739  
QY 2542 GCAGTGTGTGTTCCGAGACATGGAGAACAAAGTGTGTTGCTGAGGTGACAGCGGATGGGT 2601  
DB 1740 GCAGCCTGTGCTACGCGGACATGGAGAACAAAGTGTGTTGCGGGGATTCGGCGGGAGGGC 1799  
QY 2602 TCGTTTACGTTTGTGTTGATGACTTCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2661  
DB 1800 TGCTCTCTGTTTGTGTTGATGATTTCTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1859  
QY 2662 CTTTCTCTCAGCACCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2721  
DB 1860 CTTTCTCTCAGCACCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1919  
QY 2722 AGACAGTGTGTAACCTTCCCTGTGAGAGCCTGTGTAACCTGTTGTTGTTGTTGTTGTTGTT 2781  
DB 1920 AGACAGTGTGTAACCTTCCCTGTGAGAGCCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTT 1979  
QY 2782 TGCCTGCTCAGTCTGTTTCCCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2841  
DB 1980 TGCCGGCCCGACCGGCTATTTCCCTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2039  
QY 2842 TGTCTGTGACTACTCAGTGTATGCCAGACCTCAATTAAGACGAGGCTCACCCTTCCAGA 2901  
DB 2040 TGCAGAGGACTACTCAGTGTATGCCAGACCTCAATTAAGACGAGGCTCACCCTTCCAGA 2099  
QY 2902 GTGCTTCAAGCTGGGAGAACCATGCGTAACCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2961  
DB 2100 GCGGCTTCAAGCTGGGAGAACCATGCGTAACCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2159  
QY 2962 GTACAGGCTTATTTAGACTTGCAGGTGAACAGGCTCCAGACAGTCTGCAATCAATATAT 3021  
DB 2160 GTACAGGCTTATTTAGACTTGCAGGTGAACAGGCTCCAGACAGTCTGCAATCAATATAT 2219  
QY 3022 ACAAGATCTTCTGCTTCCAGGCTACAGGTTCCATGATGTTGTTGTTGTTGTTGTTGTTGTT 3081  
DB 2220 ACAAGATCTTCTGCTTCCAGGCTACAGGTTCCATGATGTTGTTGTTGTTGTTGTTGTTGTT 2279  
QY 3082 ACCAGGCTGTAGGAAGAACCTCAATCTTCTTGGGATCATCTCCAGGCAAGCATCCT 3141  
DB 2280 ATCAGCAAGTTTGAAGAACCCCAATTTTCTTGGGCTCATCTCTGACACGCGCTCCC 2339  
QY 3142 GCTGTATGCTATCTCTGAAGGTCAAGATCCAG 3174  
DB 2340 TCTGCTACTCTCTGAAAGCAAGCAAGCAG 2372

RESULT 14  
V72125

ID V72125 standard; cDNA; 3972 BP.

AC V72125,1999 (first entry)

DE Human catalytic telomerase sub-unit RACE-generated cDNA #4.

KW Human; catalytic telomerase subunit; therapy; diagnosis; hTC; assay;

KW modulator; treatment; inhibit; cellular disorder; death; defect; cancer;

KW ageing; antisense; neoplastic cell; telomerase-related condition; RACE;

KW tumour cell; ss.

OS Homo sapiens.

PN W09859040-A2.

PD 30-DEC-1998.

PF 09-JUN-1998; E03468.

PR 14-APR-1998; DE-016496.

PR 20-JUN-1997; DE-026329.

PR 26-MAR-1998; DE-013274.

PA (FARB ) BAYER AG.

PI Hagen G, Siegmund H, Weichel W, Wick M, Zubov D;

DR WPI; 99-081276/07.

PT New catalytically active subunit of human telomerase - used in the  
modulation of telomerase activity, particularly for treating cancer

PT and ageing  
PS Example 6; Fig 11; 76pp; German.  
CC This sequence represents a RACE-generated fragment of a novel human  
CC catalytic telomerase sub-unit (hTC). The hTC protein can be used  
CC in screening assays to identify modulators of telomerase and to treat or  
CC inhibit cellular disorders, death, defects and/or other pathological  
CC processes involving telomerase, particularly cancer and ageing (also  
CC suitable for this are agents that stimulate, inhibit or mimic the  
CC activity of the subunit). Antisense nucleic acids inhibit telomerase  
CC action (by binding to specific mRNA), particularly in neoplastic cells  
CC and may be expressed in vivo. Antibodies and fragments of the protein,  
CC used as probes or primers, are used to diagnose telomerase-related  
CC conditions (especially neoplasia) by (i) detecting abnormal levels of  
CC the subunit protein in body fluids or tissues or (ii) by measuring the  
CC amount of the encoding nucleic acid. Expression of the nucleic acid  
CC encoding the subunit mRNA is confined to tumour cells, in contrast to  
CC the ubiquitous expression of the telomerase RNA subunit.  
SQ Sequence 3972 BP; 701 A; 1287 C; 1264 G; 720 T;

Query Match 16.7%; Score 582.8; DB 1; Length 3972;  
Best Local Similarity 63.1%; Pred. No. 2.4e-157;  
Matches 1046; Conservative 0; Mismatches 522; Indels 90; Gaps 6;

QY 23 CCCGGCCTTGAGCACAATGACCCCGCTCTCTGCTTGGCCCGCGGTGCGCTCTGCTGCG 82  
DB 67 CCCGGCACCACCGGATGCGCGCGCTCCCGCTGCGGAGCCGTGCGCTCTCTGCG 126

QY 83 CAGCGATACCGGAGGTGGCGCTGGCAACCTTTGCGGGCGCTGGGGCCCGAGGG 142  
DB 127 CAGCCACTACCGGAGGTGGCGCTGGCAACCTTTGCGGGCGCTGGGGCCCGAGGG 186

QY 143 CAGCGGCTTGCAACCCGGGACCCGAGATCTACGCACTTTGGTTGGCCCAATGCCT 202  
DB 187 CTGGCGCTGTGAGCGCGGGACCGCGCGCTTCCGCGCTGTGGGCCAGTGCCT 246

QY 203 AGTGTGATGACATGGGGCTCACAGCTTCACCTGCGGACCTTCTTCCACAGGTGTC 262  
DB 247 GGTGTGCTGCGCTGGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT 306

QY 263 ATCCCTGAAGAGCTGTGGCGAGGTGTGCGAGAGCTCTGCGAGCGCAACGAGAGAA 322  
DB 307 CTGCTGAAGAGCTGTGGCGAGGTGTGCGAGAGCTGTGCGAGCGCGCGGAGAA 366

QY 323 CGTGTGCTTTTGGCTTTGAGCTTACAGCGGCGGAGCGCGGCTCCCATGGGCTT 382  
DB 367 CGTGTGCTTTGAGCTTACAGCGGCGGAGCGCGGCGCGCGCGCGCGCGCGCTT 426

QY 383 CACTAGTAGCGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 442  
DB 427 CACCAACGAGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 486

QY 443 TGCATGGATGTACTGTGAGCGGAGTGGGCGGAGCGGAGCGGAGCGGAGCGGAGCG 502  
DB 487 GCGTGTGGGCTGTGCTGTGCGCGGCTGGGCGGAGCGGAGCGGAGCGGAGCGGAG 546

QY 503 CTGTGCTTTTATCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 562  
DB 547 CTGCGCGCTTCTTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 606

QY 563 GTACCAAAATTTGCGCACCGAGATATGCGCCCTGTGCTGTGCTGTGCTGTGCTGTG 622  
DB 607 GTACCAAGCTGTGCGGCTGTGCGGCTGTGCGGCTGTGCGGCTGTGCGGCTGTGCG 666

QY 623 CGGACCGCTGGGAGGATTTTCACTAACCTTAGTGTCTTACACAGATCAAGAGCAGTAG 682  
DB 667 GCGTCTGGGATG-----GAACGGGCTGTGAACCATAG 699

QY 683 TGCACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 742  
DB 700 CGTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 759

QY 743 TCTCACCAGTACAAAGTGTGCTTACAGTAAAGAGGCGAGATGCTATCTGCTCCCGAGAGT 802

Db 760 CAGTCCAGCCGGAAGTCTGCGGTTGCCAAGAGGCCCGAGGCTGCGGCTGCCCCGTGAGCC 819

QY 803 GGAGGAGGAGCCACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 862

Db 820 GGAGGAGGAGCCGCTTGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 879

QY 863 TCCTGCTGGTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 913

Db 880 TGACCGTGGTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGT 939

QY 914 AGGAAAGGTGTGACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 970

Db 940 GGGTGGCTTCTTGGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 999

QY 971 CTCACATCTC---TGCTGTACACACCCCGGCAAAATGCTTTTACAGCTCAGGCCATTTAT 1027

Db 1000 CCCCCATCCATCGCGGCCACGAGTCCCTTGGGACGCGCTTGTCCCGCGGTGTACGC 1059

QY 1028 TGAGACCAAGATTTCTTCTTACTCCAGGAGGAGATGGCCAGAGCGTCTAAACCCCTCAT 1087

Db 1060 CGAGACCAAGCACTTCTCTTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1116

QY 1088 CCTACTCAGCAACCTCCAGCCTAACTTGACTGGGGCCAGGAGACTGGTGGAGATCATCTT 1147

Db 1117 CTTACTCAGCTCTCTGAGGCCAGCTGACTGGGCTCGGAGGCTGTGGAGACATCTT 1176

QY 1148 TCTGGGCTCAAGGCTTAGACATCAGGACCACTCTCAGAGCACACACCTCTATCGGCTG 1207

Db 1177 TCTGGGTTCCAGGCCCTGGATGCCAGGACTCCCGCAGGTTGCCCGCTGCCCGAGCG 1236

QY 1208 ATACTGGCAGATCGGCGGCTGTTCCACAGCTGTGTTGAACTGACAGAGTGCACAA 1267

Db 1237 CTACTGGCAATCGGCGGCTGTTCTTGAGCTGTGTTGAACTGACAGGAGTGCACAA 1296

QY 1268 TGTACAGCTCTCAGGTCACATTCAGAGGTTTCAACAGCAAAACACAGGTGACAGATGC 1327

Db 1297 CGGGGTGCTCTCAAGACGACTGCCCGCTGCGAGCTGCGGTCACCCAGCAGCGGTGT 1356

QY 1328 CT-----TGACACAGCC 1342

Db 1357 CTGTGCGGGAGAGCCCGAGGCTCTGTGCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAG 1416

QY 1343 ACCGCACTCATGATTTGCTCGCGCTGCACAGCTGCTGCGAGTATATGTTTCT 1402

Db 1417 CCGTGGCTGTGAGCTGCTCGCGAGCACAGCCCTGCGAGGTGTACGGCTTCT 1476

QY 1403 TCGGCGCTGTCTGCAAGGTGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1462

Db 1477 GCGGCGCTGCTGCGCGGCTGTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGG 1536

QY 1463 CCGTCTTTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAA 1522

Db 1537 CCGTCTTCTCAGGAGCAACCAAGAGTTCTCTCCCTGGGAGGAGGAGGAGGAGGAGGAGGAG 1596

QY 1523 GCAGAACTGATGTGAAGATGAAAGTAGAGATGAGGATGAGGATGAGGATGAGGATGAGGAT 1582

Db 1597 GCAGAGCTGACGTGGAAGATGAGCTGCGGAGTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1656

QY 1583 GAAGGAGGAGTGTGCGGCTGCGAGGAGCAGGCTGTGAGGG 1620

Db 1657 TGAGGAGGTGTGCGGCTGCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1694

RESULT 15  
V16979  
ID V16979 standard; DNA; 4335 BP.  
AC V16979;  
DT 13-AUG-1998 (first entry)  
DE Human telomerase reverse transcriptase genomic clone.  
KW Human: telomerase reverse transcriptase; hTERT; TRT; diagnosis;  
OS prognosis; cell proliferation; cancer; ageing; ribonucleoprotein; ss.  
OS Homo sapiens.

PN GB2317891-A.  
 PD 08-APR-1998.  
 PE 01-OCT-1997; 020890.  
 PR 14-AUG-1997; US-915503.  
 PR 01-OCT-1996; US-724643.  
 PR 18-APR-1997; US-844119.  
 PR 25-APR-1997; US-846017.  
 PR 06-MAY-1997; US-851843.  
 PR 09-MAY-1997; US-854050.  
 PR 14-AUG-1997; US-911312.  
 PR 14-AUG-1997; US-912951.  
 PA (GERO-) GERON CORP.  
 PA (UYTE-) UNIV TECHNOLOGY CORP.  
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,  
 PI Morlin GB, Nakamura T, Harley CB;  
 DR WPI: 98-171633/16.  
 PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
 PT variants - are useful in the diagnosis, prognosis and treatment of  
 PT cell proliferation conditions especially cancer and ageing  
 PS Example 3; Fig 21; 387pp; English.  
 CC The present sequence represents a human telomerase reverse transcriptase  
 CC (hTERT) genomic clone from the present invention. The present invention  
 CC also describes the following methods: (A) determining whether a test  
 CC compound is a modulator of hTERT, by detecting the change in hTERT  
 CC recombinant protein or polynucleotide, on administration of the compound;  
 CC (B) preparation of recombinant telomerase by contacting a protein  
 CC preparation of hTERT with a telomerase RNA component; (C) detection of  
 CC the hTERT RNA or protein in a sample by binding a relevant probe to the  
 CC sample and detecting the complex formed or in the case of RNA detection,  
 CC amplifying the product and correlating the presence of complex or  
 CC amplification product with presence of hTERT in the sample; and (D)  
 CC increasing the proliferation of a vertebrate cell by increasing hTERT  
 CC expression; and (E) the use of an agent that causes an increase in cell  
 CC vertebrate cell proliferation to create a medicament that inhibits  
 CC ageing. A protein preparation of hTERT and the polynucleotide encoding  
 CC hTERT can be used in the manufacture of medicaments for inhibiting the  
 CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
 CC used to treat conditions that are associated with high telomerase  
 CC activity. A protein preparation of hTERT can also be used in the new  
 CC methods.  
 SQ Sequence 4335 BP; 698 A; 1416 C; 1360 G; 861 T;

Query Match 13.4%; Score 468.8; DB 1; Length 4335;  
 Best Local Similarity 59.4%; Pred. No. 1.5e-124;  
 Matches 1046; Conservative 0; Mismatches 522; Indels 194; Gaps 7;

QY 23 CCCGCGCTTGACACATGACCGCGCTCTCGTTGCCCCCGCGTGCCTCTCTGTCGC 82  
 DB 2480 CCCGCGCACCCCGCGATGCGCGCGCTCCCGCTGCGAGCGCTCCCTGCTGCG 2539  
 QY 83 CAGCCGATACCGGAGGTGTGCGCGCTGGCAACCTTTGTGCGGCGCTGGGCGCCGAGGG 142  
 DB 2540 CAGCCACTACCGGAGGTGTGCGCGCTGGCAACCTTTGTGCGGCGCTGGGCGCCGAGGG 142  
 QY 143 CAGCGCGTGTGCAACCCCGGGGACCCGAGATCTACCGACATTTGGTTGCCCAATGCCT 202  
 DB 2600 CTGCGCGCTGTGTGAGCGCGGGGACCCGCGGCTTTCCGCGCGCTGTGTGCGCGAGTGCCT 2659  
 QY 203 AGTGTGATGACATGCGGCTACAGGCTTCCACCTGCGGACCTTTCCCTTCCAC 255  
 DB 2660 GGTGTGCGTGTGGACGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2719  
 QY 255 ----- 255  
 DB 2720 CCTCCCCGGGTGCGCGTCCGGTGGGGTTGAGGGGCGCGCGGGGAAACCGAGCATGC 2779  
 QY 255 -----CAGGTGTATCTCCCTGAAAGAGCTG 278  
 DB 2780 GGAGAGCAGCGCAGGCGACTCAGGGCGCTTCCCGCGAGGTGTCTTCCCTGAGGAGCTG 2839  
 QY 279 GTGCGCAGGGTTGTGAGAGACTCTCGAGCGCAACGAGAGAACGTGCTGGCTTTTGGC 338  
 DB 279 GTGCGCAGGGTTGTGAGAGACTCTCGAGCGCAACGAGAGAACGTGCTGGCTTTTGGC 338

DB 2840 GTGCGCCAGTGTCTGACAGAGCTGTGCGAGCGCGCGGGAAGACGTGCTGGCGCTTCGGC 2899  
 QY 339 TTTGAGCTGTCTTAACAGAGGCGAGAGCGGCGCTCCCATAGGCTTCTACTAGTAGGCTCGCT 398  
 DB 2900 TTCGCGCTGTCTGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2959  
 QY 399 AGCTACTTGGCCCAACACTGTATTATTAGAACCTTGGTGTCTAGTGTGTGATGATGCTACTG 458  
 DB 2960 AGCTACTTGGCCCAACACTGTATTATTAGAACCTTGGTGTCTAGTGTGTGATGATGCTACTG 458  
 QY 459 TTGAGCCGAGTGGCGGACGACCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCT 518  
 DB 3020 CTGCGCGCGTGGCGGACGAGCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 3079  
 QY 519 CTGCTGCCCCCAGCTGTGCTTACCAGGTGTGGGTCTCCCTGTACCAAAATTTGTGCC 578  
 DB 3080 CTGCTGCTCCAGCTGTGCTTACCAGGTGTGGCGCGCGCGCGCGCGCGCGCGCGCGCT 3139  
 QY 579 ACCACGGATATCTGGCGCTCTGTCTGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 638  
 DB 3140 GCCACTCAGCG 3197  
 QY 639 AATTTCACTAACCTTAGTGTCTTACACAGATCAAGAGCAGTAGTGTGCGGAGGAGGAGCACC 698  
 DB 3197 -----GAACGGCGCTGGAACCATAGCTCAGGAGGCGCGG 3232  
 QY 699 AAACCCCTGCGCTTGCATCTCAGAGTACAAAGAGGATCTGAGTCTCACCAGTACAAGT 758  
 DB 3233 GTCCCTTGGCGCTGCG 3292  
 QY 759 GTCCCTTACGCTAAGAGGCGCAGATGCTATCTGTCCGAGAGTGTGAGGAGGAGGAGCCAC 818  
 DB 3293 CTGCGGTGCGCAAGAGGCGCGCGCTGCGCTGCGCTTACGCGGAGGCGGAGCGCGCTT 3352  
 QY 819 AGCAGGTGCTTACCAACCCGATCAGCAATCATGGGTGCCAAGTCTCTCTCGTCCCGCC 878  
 DB 3353 GGCAGGGGTCTTGGCG 3412  
 QY 879 GAGTGCCTTACT-----GCAGAGAAAGATTGTCTTCTTAAGGAAGGTGTCTGAC 929  
 DB 3413 GTGCTGCTACCTGCCAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTGCG 3472  
 QY 930 CTGAGTCTCTC---TGGTGGTGTGTCTTAACACACAGCCAGCTCCACAT---CTCTG 983  
 DB 3473 ACAGCGCACTTCCACCCATCCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3532  
 QY 984 CTGTACCAACCGCGCGCAAAATGCTTTCAGCTCAGGCGCATTTATTGAGACCAAGATTTTC 1043  
 DB 3533 CGGCCACCACTGCTTGGGACACGCGCTTGTCCCGCGGTGTAGCGCGGAGACCAAGCACATTC 3592  
 QY 1044 CTTTACTCCAGGGGAGATGGCCCAAGAGCGTCTAAACCCCTCATCTCTCTCTCTCTCTCT 1103  
 DB 3593 CTCTACTCTCAGGCGA---CAAGGAGCAGTGTGCGCGCGCTCTCTCTCTCTCTCTCTCT 3649  
 QY 1104 CAGCTTAAGTCTGAGTGGGCGCAGGAGTGTGTGAGATCATCTTTCTGGGCTCAAGGCT 1163  
 DB 3650 AGGCGCGCGCTGACTTGGGCTGCGAGGCTGTGTGAGAGCATCTTTCTGGTTCAGGCGCC 3709  
 QY 1164 AGGACATCAGGACCACTCTGCGAGGACACACCGTCTATCGCGTCTGATCTGCGAGTGGG 1293  
 DB 3710 TGATGCGCAGGACCTCCCGCGAGGTGTCCCGCGCTGTGCGCGCGCTTGTGCGCAATGCGG 3769  
 QY 1224 CCGCTGTTCACAGCTCTGCTGTGAACCATGCAGATGCCAATATGTCTGAGATCTCTCAGG 1283  
 DB 3770 CCGCTGTTCAGGAGCTGCTTGGGAACCCAGCGAGTGTGCGCGCTCTCTCTCTCTCTCT 3829  
 QY 1284 TCACATTCCAGGTTTCGAACAGCAACCAACAGGTGTGAGATGCTT----- 1331  
 DB 3830 AGGCACTGCGCGCTGCGAGCTGCGGTACCCCGCAGCAGCGCGGTGTGTGTGCGCGGAGAG 3889  
 QY 1331 -----GAACACCG 1358  
 DB 3890 CCGCAGGCGCTGTGTGGCG 3949

Mon Aug 16 10:21:24 1999

QY 1359 TTGCTCCGCTGCACAGAGTCCCTGGCAGGTATATGGTTTTCTTGGGCGCTGTCTGTC 1418  
 DB 3950 CTGCTCCGCGACAGACAGAGCCCTGGCAGGTGTACGGCTTCGTGCGGGCGCTGCTGCGC 4009  
 QY 1419 AAGGTGGTGTCTGTAGTCTCTGGGGTACCAGGCACAAATGAGCGCGCTTCTTTAAGAAC 1478  
 DB 4010 CGGCTGGTGGCCCCCAGGCTCTGGGGCTCCAGGCACACAGAACGCCGCTTCTCAGGAAC 4069  
 QY 1479 TTAAGAAGTTTCATCTCGTTGGGGAATACGGCAAGCTATCACTGCAGGAACGTATGTGG 1538  
 DB 4070 ACCAAGAAGTTTCATCTCCCTGGGGAAGCATGCCAAGCTCTGCTGCAGGAGCTGACGTGG 4129  
 QY 1539 AAGATGAAGTAGAGGATGCCACTGGCTCCGCGCAGCGCGGGAAGACCGTGTCCCC 1598  
 DB 4130 AAGATGAGCGTCGGGAGTGGCGCTTGGCTGCGCAGGAGCCAGGTGAGGAGGTGGTGCC 4189  
 QY 1599 GCTGCAGAGCACCGTCTGAGGG 1620  
 DB 4190 CTCGAGGGCCCCAGGCCCCAGAG 4211

Search completed: August 13, 1999, 13:03:09  
 Job time: 6142 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 1999, 21:18:14 ; Search time 17.01 seconds  
(without alignments)  
1562.364 Million cell updates/sec

Title: US-09-042-460-2  
Perfect score: 5901  
Sequence: 1 MTRAPRCVAVRSLRLSRVRE.....TILKAAADPALSTDFQTILD 1122

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database: A\_Geneseq\_36.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3505	59.4	1189	1 W47008	Glutathione-S-tran
2	3505	59.4	1132	1 W46957	Human telomerase r
3	3505	59.4	1154	1 W61350	Human telomerase p
4	3505	59.4	1132	1 W90251	Human catalytic te
5	3499	59.3	1285	1 W47000	HIS tagged thiores
6	3498	59.3	1132	1 W71376	Human telomerase c
7	3496	59.2	1132	1 W56113	Human telomerase r
8	3486.5	59.1	1199	1 W47007	Glutathione-S-tran
9	3463	58.7	1405	1 W56101	Enhanced green flu
10	3320	56.3	1150	1 W47005	Glutathione-S-tran
11	2790.5	47.3	949	1 W61349	Human telomerase p
12	2195	37.2	807	1 W46997	Human telomerase r
13	2073	35.1	591	1 W37384	A catalytic telome
14	1506.5	25.5	564	1 W56109	Human telomerase r
15	1092.5	18.5	538	1 W47001	Glutathione-S-tran
16	941	15.9	531	1 W47002	Glutathione-S-tran
17	861	14.6	514	1 W47003	Glutathione-S-tran
18	758.5	12.9	514	1 W47004	Glutathione-S-tran
19	736	12.5	259	1 W46998	Human telomerase r
20	714	12.1	516	1 W47005	Glutathione-S-tran
21	621.5	10.5	988	1 W56107	S. pombe trt prote
22	465	7.9	131	1 W37385	Amino acid sequenc
23	410.5	7.0	1031	1 W56102	Euplotes telomeras
24	341	5.8	884	1 W59778	Amino acid sequenc
25	341	5.8	884	1 W71375	Yeast telomerase c
26	136	2.3	30	1 W57395	Human telomerase r
27	131	2.2	778	1 R94900	Human tasin. Mamm
28	126	2.1	24	1 W57391	Human telomerase r
29	126	2.1	24	1 W57387	Human telomerase r
30	115	1.9	28	1 W57394	Human telomerase r
31	110	1.9	1239	1 R45945	Glutamic acid rece
32	106	1.8	1364	1 R94345	Hepatitis GB virus
33	106	1.8	1349	1 W59359	Human retinal dege
34	105.5	1.8	876	1 W59032	B. pallidus DNA po
35	105.5	1.8	1084	1 W59033	B. pallidus DNA po
36	105.5	1.8	1276	1 W59034	B. pallidus DNA po
37	102.5	1.7	566	1 W82559	A. thaliana homolo
38	102	1.7	1686	1 W70991	Human class II p13
39	101.5	1.7	645	1 W42084	Amino acid sequenc
40	101	1.7	677	1 W06083	Human TATA-binding
41	101	1.7	677	1 W25017	TATA-binding prote
42	99	1.7	1219	1 R66063	Human NMDAR2 recep
43	99	1.7	1214	1 R66065	Human NMDAR2 recep

## ALIGNMENTS

RESULT 1	44	99	1.7	1621	1	W31228	Human Immunodefici
W47008	45	99	1.7	158	1	W46999	Tetrahymena macron
ID W47008 standard; Protein; 1189 AA.							
AC W47008:							
DT 13-AUG-1998 (first entry)							
DE Glutathione-S-transferase and hTERT fusion protein 8.							
KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;							
KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein.							
OS Synthetic.							
OS Homo sapiens.							
FH Key							
FT Misc_difference 22..23							
FT /note= "enterokinase cleavage site"							
PN GB2317891-A.							
PD 08-APR-1998.							
PF 01-OCT-1997; 020890.							
PR 14-AUG-1997; US-915503.							
PR 01-OCT-1996; US-724643.							
PR 18-APR-1997; US-844419.							
PR 25-APR-1997; US-846017.							
PR 06-MAY-1997; US-851843.							
PR 09-MAY-1997; US-854050.							
PR 14-AUG-1997; US-911312.							
PR 14-AUG-1997; US-912951.							
FA (GERO-) GERON CORP.							
FA (UYTE-) UNIV TECHNOLOGY CORP.							
PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,							
PI Morin GB, Nakamura T, Harley CB;							
DR WPI: 98-171633/16.							
PT Pure and recombinant human Telomerase Reverse Transcriptase and its							
PT variants - are useful in the diagnosis, prognosis and treatment of							
PT cell proliferation conditions especially cancer and ageing							
PS Example 6; Page 234-235; 387pp; English.							
CC The present sequence represents a fusion protein from an example							
CC of the present invention which describes human telomerase reverse							
CC transcriptase (hTERT). The present invention also describes the							
CC following methods: (A) determining whether a test compound is							
CC a modulator of hTERT, by detecting the change in hTERT recombinant							
CC protein or polynucleotide, on administration of the compound;							
CC (B) preparation of recombinant telomerase by contacting a protein							
CC preparation of hTERT with a telomerase RNA component; (C) detection							
CC of the hTERT RNA or protein in a sample by binding a relevant							
CC probe to the sample and detecting the complex formed or in the case of							
CC RNA detection, amplifying the product and correlating the presence of							
CC complex or amplification product with presence of hTERT in the sample;							
CC and (D) increasing the proliferation of a vertebrate cell by increasing							
CC hTERT expression; and (E) the use of an agent that causes an increase in							
CC cell vertebrate cell proliferation to create a medication that inhibits							
CC ageing. A protein preparation of hTERT and the polynucleotide encoding							
CC hTERT can be used in the manufacture of medicaments for inhibiting the							
CC effect of ageing or cancer. Inhibitors of telomerase activity can be							
CC used to treat conditions that are associated with high telomerase							
CC activity. A protein preparation of hTERT can also be used in the new							
CC methods.							
CC Sequence 1189 AA;							

Query Match	59.4%	Score 3505;	DB 1;	Length 1189;
Best Local Similarity	62.4%	Pred. NO. 0;		
Matches 719;	Conservative 122;	Mismatches 260;	Indels 52;	Gaps 13;
QY 1	MTRAPRCVAVRSLRLSRVREVP	PLATFVRRLGPGRRLVQDPGPKYRTLV	AOCLVCMHW 60	
Db 58	MPRAPRCVAVRSLRLSRVREVP	PLATFVRRLGPGRRLVQDPGPKYRTLV	AOCLVCMHW 117	
QY 61	GSQPPADLSFHQVSSLSKELV	ARVVRVQRLCERNERNVLA	FGFELLNEARGGPPMAFTSSVR 120	





```

Db 181 ATQARPPPHAS -GPRRLG-----CERAWNSVREAGVPLGLPAPGARRRGGSSASRS 231
QY 241 VPSAKKARCYVPVREBP-----HRQVLPSPGKSW-VPSPARSPPEVPTAEKDLSSK 292
Db 232 LPLKPRRGRGAPEPERTPVCGSWAHPRGRGSDRGFCVVSAPAR-----PAEATSLE 286
QY 293 GKVSLSLS-GSVCKKHPSTSLSPRONAFQLRP-FIETRHFLYSRGDGERLNPSF 350
Db 287 GALSOTRSHSPVGRQHAGPPSTSRPRPDWTPCPVYAEATHFLYSSGD-KEQLRPSF 345
QY 351 LLSNLOPLNLTGARLVEIFLGSRPRTSGPLCTRHSRRYQWRPLFOQLLVNHAECQY 410
Db 346 LLSLRPSLTGARLVETIFLGSRPMPGTPRRLPLPQRYQWRPLFLELLGNHAQCPY 405
QY 411 VRLSRHCRTANQQVTDAL-----NTSPHMLDLRLHSSPWQY 452
Db 406 GVLLKTHCPRAA---VTPAAGVCAKPKGWSVAAPPEEDTPRRLVQLLRHSSPWQY 462
QY 453 GFLRACLCKVVSASLWCTRHNERRFFKNLKKFISLKGKGLSLQELMWMKMYEDCHWLSR 512
Db 463 GFVRACLRLRVLPPCLMGSRHNERFLNRTKFFISLKGKHAQLSLQELTWKMSVRDCAWLR 522
QY 513 SPGRDVPAAEHLRERILATFLWMDTVVOLLRSFFYITESTFQKNLFYFRKSVMS 572
Db 523 SPGVGCVPAAEHLRERILAKFLHMLMSVYVVELLSRFFYVETTFQKNLFYFRKSVMS 582
QY 573 KLOSIGVQHLRVLRELSEAEVRHODTWLAMPICRLRFPKPNGLRPVNMYSMT 632
Db 583 KLOSIGVQHLRVLRELSEAEVRHODTWLAMPICRLRFPKPNGLRPVNMYSMT 632
QY 633 RALGRKQAHQFQRTUKTLFMSLNVERTKPHLMGSSVLGMNDIYTRAFVLRVRLDQ 692
Db 643 RTRERKRAELTSRYKALFSLVNLVNERARRPGLGASVGLDDIHRNRTFVLRVRAQDP 702
QY 693 TPRIYVKADVTGAYDAIPOGKLVEVANNIRHSESTCIROYAVVRDSDQVHKSPRR 752
Db 703 PPELYPKVDVTGAYDTIPQDRLETVIASIIR-PONTYCVRRYAVVQAAHGRVKAFKS 761
QY 753 QVTTLSLQPYMGQFLKHLQDSASALRNSVVIEQISIMNNESSLSLDFELHFLRHSVVK 812
Db 762 HVSTLTLDQPYMRQFVAHLQET--SPLRDVAVIEQSSSLNEASSGLDFVFRFCHHAVER 819
QY 813 IGDRCYTCOGTPOGSSLSLCLSCFGDMENKLFPAEQVRDGLLRLFRVDDFLVTPHLDQ 872
Db 820 IRGKSYVOCQGTPOGSSLSLCLSCYGDENKLFAGIRRDGLLRLVDDFLVTPHLDQ 879
QY 873 AKTFSTLVHGVPEYGCMLNLTQVNFPEVPGTLGGAPYQLPAHCLFPWCGLLDTQT 932
Db 880 AKTFSTLVHGVPEYGCMLNLTQVNFPEVPGTLGGAPYQLPAHCLFPWCGLLDTQT 932
QY 933 LEVCDYSYGAQTSKTSLSLTSQVFKAGKTMNRKLLSVLRKCHGLFLDLQVNSLQTVCI 992
Db 940 LEVCDYSYGAQTSKTSLSLTSQVFKAGKTMNRKLLSVLRKCHGLFLDLQVNSLQTVCI 999
QY 993 NIYKIFLQAYRFHACVQLPDDQVRKNTLFFGLGISSOASCCYAILKVNPGMTLKAS 1052
Db 1000 NIYKIFLQAYRFHACVQLPDDQVRKNTLFFGLGISSOASCCYAILKVNPGMTLKAS 1052
QY 1053 GS---FPEAAHLCYQAFLLKLAHSYIYKCLGLPRTAQKLLCRKLPEATWILKAAA 1109
Db 1060 GAAGPLSEAVQWLCHQAFLLKLRHRTVYVPLGLSLRTAQKLLCRKLPEATWILKAAA 1109
QY 1110 DPALSTDFQITLD 1122
Db 1120 NPALPSDFKITLD 1132

```

RESULT 3

W61350

ID W61350 standard; Protein: 1154 AA.

AC W61350;

DT 12-OCT-1998 (first entry)

```

DE Human telomerase protein 2 (TP2).
KW TP2; human; telomerase protein 2; cancer; AIDS; ageing; therapy.
OS Homo sapiens.
PN W09821343-A1.
PD 22-MAY-1998.
PF 13-NOV-1997; U21248.
PR 16-OCT-1997; US-951733.
PR 15-NOV-1996; US-871189.
PR 11-JUN-1997; US-873039.
PA (AMGE-) AMGEN CANADA INC.
PI (AMGE-) AMGEN INC.
PI Harrington LA, Robinson MO;
DR WPI: 98-297946/26.
DR N-PSDB: V27876.
PT New nucleic acid encoding human telomerase protein-2 - used for
PT regulating telomerase activity, e.g. for treating cancer or acquired
PT immune deficiency syndrome
PS Claim 1e; Fig 9; 150pp; English.
CC This polypeptide comprises human telomerase protein 2 (TP2), a
CC novel protein of the telomerase complex. Its amino acid sequence
CC was deduced from a composite (see V27876) of isolated cDNA clones
CC 32 (see V27872) and TP2-15 (see V27875), obtained from a human
CC colon tumour cell line LIM1863 cDNA. Expressing TP2 in a cell is
CC used to increase telomerase activity and thus proliferation for
CC treatment of e.g. HIV infection, AIDS and ageing disorders, while
CC expressing an inactive mutant of TP2 (or molecule antisense to the
CC gene) is used to decrease telomerase activity, e.g. for treatment
CC of cancer. TP2 polypeptides can also be used to screen for agents
CC that inhibit TP2 activity or its binding to TRIP1 (see W61347) or
CC telomerase RNA, potentially useful therapeutically, also to raise
CC specific antibodies useful in immunoassays and therapeutically as
CC inhibitors. Also contemplated are transgenic animals in which the
CC TP2 gene has been inactivated or is overexpressed. TP2 polypeptides
CC are administered i.v., s.c. or orally, or they are delivered from
CC engineered cells or gene therapy vectors.
SQ Sequence 1154 AA;

```

```

Query Match 59.4%; Score 3505; DB 1; Length 1154;
Best Local Similarity 62.4%; Pred. No. 0;
Matches 719; Conservative 122; Mismatches 260; Indels 52; Gaps 13;

```

```

QY 1 MTRAPCPAVRSLRSRVREWVPLATFVRLRGLPEGRRLVQPGDKPIRYTLVAQCLVCMHW 60
Db 23 MTRAPCPAVRSLRSRVREWVPLATFVRLRGLPEGRRLVQPGDKPIRYTLVAQCLVCMHW 60
QY 61 GSQPPADLSHFQVSLKELVARVQRLCERNERNVLAFLGELLNEARGGPPMAFTSSVR 120
Db 83 DARPPAAPSPFQVSLKELVARVQRLCERNERNVLAFLGELLNEARGGPPMAFTSSVR 142
QY 121 SYLPNTVETLRVSGAMMLLSRVGDDLLVYLLAHALCALVLLVPPSCAYOVCGSPLYOICA 180
Db 143 SYLPNTVETLRVSGAMMLLSRVGDDLLVYLLAHALCALVLLVPPSCAYOVCGSPLYOICA 202
QY 181 TTDIWPVSASVRYTRPVGRNFTNLRLFOQLKSSRQEAQKPLALPSRGTKRHLSTSTS 240
Db 203 ATQARPPPHAS -GPRRLG-----CERAWNSVREAGVPLGLPAPGARRRGGSSASRS 253
QY 241 VPSAKKARCYVPVREBP-----HRQVLPSPGKSW-VPSPARSPPEVPTAEKDLSSK 292
Db 254 LPLKPRRGRGAPEPERTPVCGSWAHPRGRGSDRGFCVVSAPAR-----PAEATSLE 308
QY 293 GKVSLSLS-GSVCKKHPSTSLSPRONAFQLRP-FIETRHFLYSRGDGERLNPSF 350
Db 309 GALSOTRSHSPVGRQHAGPPSTSRPRPDWTPCPVYAEATHFLYSSGD-KEQLRPSF 367
QY 351 LLSNLOPLNLTGARLVEIFLGSRPRTSGPLCTRHSRRYQWRPLFOQLLVNHAECQY 410
Db 368 LLSLRPSLTGARLVETIFLGSRPMPGTPRRLPLPQRYQWRPLFLELLGNHAQCPY 427
QY 411 VRLSRHCRTANQQVTDAL-----NTSPHMLDLRLHSSPWQY 452
Db 428 GVLLKTHCPRAA---VTPAAGVCAKPKGWSVAAPPEEDTPRRLVQLLRHSSPWQY 484

```

Mon Aug 16 10:21:26 1999

PT	and ageing	PT	Claim 2: Fig 2: 76pp: German.	PT	and ageing
PS	This sequence represents a novel human catalytic telomerase sub-unit	PS	This sequence represents a novel human catalytic telomerase sub-unit	PS	This sequence represents a novel human catalytic telomerase sub-unit
CC	(hlc). This protein can be used in screening assays to identify	CC	(hlc). This protein can be used in screening assays to identify	CC	(hlc). This protein can be used in screening assays to identify
CC	modulators of telomerase and to treat or inhibit cellular disorders,	CC	modulators of telomerase and to treat or inhibit cellular disorders,	CC	modulators of telomerase and to treat or inhibit cellular disorders,
CC	death, defects and/or other pathological processes involving telomerase,	CC	death, defects and/or other pathological processes involving telomerase,	CC	death, defects and/or other pathological processes involving telomerase,
CC	particularly cancer and ageing (also suitable for this agents that	CC	particularly cancer and ageing (also suitable for this agents that	CC	particularly cancer and ageing (also suitable for this agents that
CC	stimulate, inhibit or mimic the activity of the subunit). Antisense	CC	stimulate, inhibit or mimic the activity of the subunit). Antisense	CC	stimulate, inhibit or mimic the activity of the subunit). Antisense
CC	nucleic acids inhibit telomerase action (by binding to specific mRNA),	CC	nucleic acids inhibit telomerase action (by binding to specific mRNA),	CC	nucleic acids inhibit telomerase action (by binding to specific mRNA),
CC	particularly in neoplastic cells and may be expressed in vivo. Antibodies	CC	particularly in neoplastic cells and may be expressed in vivo. Antibodies	CC	particularly in neoplastic cells and may be expressed in vivo. Antibodies
CC	and fragments of the protein, used as probes or primers, are used to	CC	and fragments of the protein, used as probes or primers, are used to	CC	and fragments of the protein, used as probes or primers, are used to
CC	diagnose telomerase-related conditions (especially neoplasia) by (i)	CC	diagnose telomerase-related conditions (especially neoplasia) by (i)	CC	diagnose telomerase-related conditions (especially neoplasia) by (i)
CC	detecting abnormal levels of the subunit protein in body fluids or	CC	detecting abnormal levels of the subunit protein in body fluids or	CC	detecting abnormal levels of the subunit protein in body fluids or
CC	tissues or (ii) by measuring the amount of the encoding nucleic acid.	CC	tissues or (ii) by measuring the amount of the encoding nucleic acid.	CC	tissues or (ii) by measuring the amount of the encoding nucleic acid.
CC	Expression of the nucleic acid encoding the subunit mRNA is confined to	CC	Expression of the nucleic acid encoding the subunit mRNA is confined to	CC	Expression of the nucleic acid encoding the subunit mRNA is confined to
CC	tumour cells, in contrast to the ubiquitous expression of the telomerase	CC	tumour cells, in contrast to the ubiquitous expression of the telomerase	CC	tumour cells, in contrast to the ubiquitous expression of the telomerase
CC	RNA subunit.	CC	RNA subunit.	CC	RNA subunit.
CC	Sequence 1132 AA:	CC	Sequence 1132 AA:	CC	Sequence 1132 AA:
QY	1 MTRAPRCVRSLLRSRYREVWPLATFVRRLGPEGRLVQPGDPKTYRTLVACQLVCMHW 60	QY	1 MTRAPRCVRSLLRSRYREVWPLATFVRRLGPEGRLVQPGDPKTYRTLVACQLVCMHW 60	QY	1 MTRAPRCVRSLLRSRYREVWPLATFVRRLGPEGRLVQPGDPKTYRTLVACQLVCMHW 60
DB	1 MPAPRCVRSLLRSRYREVWPLATFVRRLGPEGRLVQPGDPKTYRTLVACQLVCMHW 60	DB	1 MPAPRCVRSLLRSRYREVWPLATFVRRLGPEGRLVQPGDPKTYRTLVACQLVCMHW 60	DB	1 MPAPRCVRSLLRSRYREVWPLATFVRRLGPEGRLVQPGDPKTYRTLVACQLVCMHW 60
QY	61 GSOPPPADLSFHQVSSLLKELVARVQRLCRNERNVLAFTGELLNEARGPPMAFTSSVR 120	QY	61 GSOPPPADLSFHQVSSLLKELVARVQRLCRNERNVLAFTGELLNEARGPPMAFTSSVR 120	QY	61 GSOPPPADLSFHQVSSLLKELVARVQRLCRNERNVLAFTGELLNEARGPPMAFTSSVR 120
DB	61 DARPPAPSPFQVSCLELVARVQRLCRNERNVLAFTGELLNEARGPPMAFTSSVR 120	DB	61 DARPPAPSPFQVSCLELVARVQRLCRNERNVLAFTGELLNEARGPPMAFTSSVR 120	DB	61 DARPPAPSPFQVSCLELVARVQRLCRNERNVLAFTGELLNEARGPPMAFTSSVR 120
QY	121 SYLNTVIEILVSGANMLLSRVGGDLLVYLLAHALYLLVPPSCAYQVCGSPLXOICA 180	QY	121 SYLNTVIEILVSGANMLLSRVGGDLLVYLLAHALYLLVPPSCAYQVCGSPLXOICA 180	QY	121 SYLNTVIEILVSGANMLLSRVGGDLLVYLLAHALYLLVPPSCAYQVCGSPLXOICA 180
DB	121 SYLNTVTDALRGSGANMLLSRVGGDLLVYLLAHALYLLVPPSCAYQVCGSPLXOICA 180	DB	121 SYLNTVTDALRGSGANMLLSRVGGDLLVYLLAHALYLLVPPSCAYQVCGSPLXOICA 180	DB	121 SYLNTVTDALRGSGANMLLSRVGGDLLVYLLAHALYLLVPPSCAYQVCGSPLXOICA 180
QY	181 TTDWPSVASYRTPRVGRNFTNLRFLOQIKSSROEAPKPLALSRGTRKHLSTSTS 240	QY	181 TTDWPSVASYRTPRVGRNFTNLRFLOQIKSSROEAPKPLALSRGTRKHLSTSTS 240	QY	181 TTDWPSVASYRTPRVGRNFTNLRFLOQIKSSROEAPKPLALSRGTRKHLSTSTS 240
DB	181 ATQAPPPHAS-GPRLRLG-----CERAWNHSVREAGVPLGAPAGARRGGSSASRS 231	DB	181 ATQAPPPHAS-GPRLRLG-----CERAWNHSVREAGVPLGAPAGARRGGSSASRS 231	DB	181 ATQAPPPHAS-GPRLRLG-----CERAWNHSVREAGVPLGAPAGARRGGSSASRS 231
QY	241 VSAKAKACYPVPRVEEGP-----HRQVLPSPSGSW-VPSPARSPVPTAEKDLSSK 292	QY	241 VSAKAKACYPVPRVEEGP-----HRQVLPSPSGSW-VPSPARSPVPTAEKDLSSK 292	QY	241 VSAKAKACYPVPRVEEGP-----HRQVLPSPSGSW-VPSPARSPVPTAEKDLSSK 292
DB	232 LPLPKRPRGAAPERTFVGQSWAHGPRTRGSDRGFCVVSFAR-----PAEATFLE 286	DB	232 LPLPKRPRGAAPERTFVGQSWAHGPRTRGSDRGFCVVSFAR-----PAEATFLE 286	DB	232 LPLPKRPRGAAPERTFVGQSWAHGPRTRGSDRGFCVVSFAR-----PAEATFLE 286
QY	293 GRVSDLSL-GSVCKKHKPSSTSLSPQNAFOLRP-FIETRHFLSRGDCQERLPSF 350	QY	293 GRVSDLSL-GSVCKKHKPSSTSLSPQNAFOLRP-FIETRHFLSRGDCQERLPSF 350	QY	293 GRVSDLSL-GSVCKKHKPSSTSLSPQNAFOLRP-FIETRHFLSRGDCQERLPSF 350
DB	287 GALSCTRSHSPVGRQHHAGPPSTSRPPRPMWDTCPVPYATKFKLYSSGD-KEQLRPSF 345	DB	287 GALSCTRSHSPVGRQHHAGPPSTSRPPRPMWDTCPVPYATKFKLYSSGD-KEQLRPSF 345	DB	287 GALSCTRSHSPVGRQHHAGPPSTSRPPRPMWDTCPVPYATKFKLYSSGD-KEQLRPSF 345
QY	351 LLSNLOPNLTGARRLVEIFLGSRRPRTSGPLCRTHRLSRRYWQMRPLFOQLLVNHAECQY 410	QY	351 LLSNLOPNLTGARRLVEIFLGSRRPRTSGPLCRTHRLSRRYWQMRPLFOQLLVNHAECQY 410	QY	351 LLSNLOPNLTGARRLVEIFLGSRRPRTSGPLCRTHRLSRRYWQMRPLFOQLLVNHAECQY 410
DB	346 LLSLRPSLTGARRLVETIFLGSRRPMPPTPRPLPRLPQRYWQMRPLFLELLGNHAOCY 405	DB	346 LLSLRPSLTGARRLVETIFLGSRRPMPPTPRPLPRLPQRYWQMRPLFLELLGNHAOCY 405	DB	346 LLSLRPSLTGARRLVETIFLGSRRPMPPTPRPLPRLPQRYWQMRPLFLELLGNHAOCY 405
QY	411 VRLSRHCRFTANOQVTDAL-----NTSPHMLDLRLHSSPWQY 452	QY	411 VRLSRHCRFTANOQVTDAL-----NTSPHMLDLRLHSSPWQY 452	QY	411 VRLSRHCRFTANOQVTDAL-----NTSPHMLDLRLHSSPWQY 452
DB	406 GYLLKTHCPRLAA---VTPAAGVCAREKPOGSVAAPEDDTPRLVQLLRHSSPWQY 462	DB	406 GYLLKTHCPRLAA---VTPAAGVCAREKPOGSVAAPEDDTPRLVQLLRHSSPWQY 462	DB	406 GYLLKTHCPRLAA---VTPAAGVCAREKPOGSVAAPEDDTPRLVQLLRHSSPWQY 462
QY	453 GFRLACLKLVASLNGTRHNRERFKNLKFISLGKYGKLSLQELMKMKVEDCHWLSR 512	QY	453 GFRLACLKLVASLNGTRHNRERFKNLKFISLGKYGKLSLQELMKMKVEDCHWLSR 512	QY	453 GFRLACLKLVASLNGTRHNRERFKNLKFISLGKYGKLSLQELMKMKVEDCHWLSR 512
DB	463 GFVRACLRLVPPGLWGSRRHNRERFKNLKFISLGKYGKLSLQELMKMKVEDCHWLSR 522	DB	463 GFVRACLRLVPPGLWGSRRHNRERFKNLKFISLGKYGKLSLQELMKMKVEDCHWLSR 522	DB	463 GFVRACLRLVPPGLWGSRRHNRERFKNLKFISLGKYGKLSLQELMKMKVEDCHWLSR 522
QY	513 SPCKDRVPAEHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 572	QY	513 SPCKDRVPAEHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 572	QY	513 SPCKDRVPAEHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 572
DB	523 SPQGVCPAAEHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 582	DB	523 SPQGVCPAAEHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 582	DB	523 SPQGVCPAAEHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 582
QY	573 KLOSIGVROHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 632	QY	573 KLOSIGVROHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 632	QY	573 KLOSIGVROHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 632
DB	583 KLOSIGVROHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 642	DB	583 KLOSIGVROHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 642	DB	583 KLOSIGVROHLRERILATLFLWMDTYVYVQLLSRFFYITESTFQXNRFLFFYKKSWS 642
QY	633 RALGRKQAOHFTORLKTFLSMNLNERTKHPHLMGSSVLGMNDIYKTRAFVLRVRLDQ 692	QY	633 RALGRKQAOHFTORLKTFLSMNLNERTKHPHLMGSSVLGMNDIYKTRAFVLRVRLDQ 692	QY	633 RALGRKQAOHFTORLKTFLSMNLNERTKHPHLMGSSVLGMNDIYKTRAFVLRVRLDQ 692
DB	643 RTFREKRAELTSRVKALFSLVNLNERTKHPHLMGSSVLGMNDIYKTRAFVLRVRLDQ 702	DB	643 RTFREKRAELTSRVKALFSLVNLNERTKHPHLMGSSVLGMNDIYKTRAFVLRVRLDQ 702	DB	643 RTFREKRAELTSRVKALFSLVNLNERTKHPHLMGSSVLGMNDIYKTRAFVLRVRLDQ 702
QY	693 TPRMYFVADVTGAYDAIPQGLKLVAVVWNRHSESTYCIROYAVVRRDSOGQVHKSFR 752	QY	693 TPRMYFVADVTGAYDAIPQGLKLVAVVWNRHSESTYCIROYAVVRRDSOGQVHKSFR 752	QY	693 TPRMYFVADVTGAYDAIPQGLKLVAVVWNRHSESTYCIROYAVVRRDSOGQVHKSFR 752
DB	725 PPLEYFVKVDVTGAYDTIPQDRLTEVIAIILK-PQNTYCVRYAVVQAAHGHVRKAFKS 783	DB	725 PPLEYFVKVDVTGAYDTIPQDRLTEVIAIILK-PQNTYCVRYAVVQAAHGHVRKAFKS 783	DB	725 PPLEYFVKVDVTGAYDTIPQDRLTEVIAIILK-PQNTYCVRYAVVQAAHGHVRKAFKS 783
QY	753 QVTTLSLQPYMGOFKHLQDSDASALRNSVVEIOSISNMSSSLDFDFFLHFLRHSVVK 812	QY	753 QVTTLSLQPYMGOFKHLQDSDASALRNSVVEIOSISNMSSSLDFDFFLHFLRHSVVK 812	QY	753 QVTTLSLQPYMGOFKHLQDSDASALRNSVVEIOSISNMSSSLDFDFFLHFLRHSVVK 812
DB	784 HVTSLDLPYMQQFVAHLQET--SPLRDVAVIEQSSSLNEASSGLDFVFLRFMCHHAVR 841	DB	784 HVTSLDLPYMQQFVAHLQET--SPLRDVAVIEQSSSLNEASSGLDFVFLRFMCHHAVR 841	DB	784 HVTSLDLPYMQQFVAHLQET--SPLRDVAVIEQSSSLNEASSGLDFVFLRFMCHHAVR 841
QY	813 IGDRCTYQCGIPQGSLSLTLSCFCGDMENKLFAYVORDGLLRFVDDFLVTPHLDO 872	QY	813 IGDRCTYQCGIPQGSLSLTLSCFCGDMENKLFAYVORDGLLRFVDDFLVTPHLDO 872	QY	813 IGDRCTYQCGIPQGSLSLTLSCFCGDMENKLFAYVORDGLLRFVDDFLVTPHLDO 872
DB	842 IRGKSYVQCQIPQGSLSLTLSCFCGDMENKLFAYVORDGLLRFVDDFLVTPHLDO 901	DB	842 IRGKSYVQCQIPQGSLSLTLSCFCGDMENKLFAYVORDGLLRFVDDFLVTPHLDO 901	DB	842 IRGKSYVQCQIPQGSLSLTLSCFCGDMENKLFAYVORDGLLRFVDDFLVTPHLDO 901
QY	873 AKTFLSLVGVPEYGMINLQTVNVPVEPGTGLGAAPYQALPCHLPFWCGLLDQ 932	QY	873 AKTFLSLVGVPEYGMINLQTVNVPVEPGTGLGAAPYQALPCHLPFWCGLLDQ 932	QY	873 AKTFLSLVGVPEYGMINLQTVNVPVEPGTGLGAAPYQALPCHLPFWCGLLDQ 932
DB	902 AKTFLSLVGVPEYGMINLQTVNVPVEPGTGLGAAPYQALPCHLPFWCGLLDQ 961	DB	902 AKTFLSLVGVPEYGMINLQTVNVPVEPGTGLGAAPYQALPCHLPFWCGLLDQ 961	DB	902 AKTFLSLVGVPEYGMINLQTVNVPVEPGTGLGAAPYQALPCHLPFWCGLLDQ 961
QY	933 LEVFCDSYGAQTSIKTSLTFQSVFAGKTMRNKLLSVLRKCHGLFLDLOVNSLQTVCI 992	QY	933 LEVFCDSYGAQTSIKTSLTFQSVFAGKTMRNKLLSVLRKCHGLFLDLOVNSLQTVCI 992	QY	933 LEVFCDSYGAQTSIKTSLTFQSVFAGKTMRNKLLSVLRKCHGLFLDLOVNSLQTVCI 992
DB	962 LEVFCDSYGAQTSIKTSLTFQSVFAGKTMRNKLLSVLRKCHGLFLDLOVNSLQTVCI 1021	DB	962 LEVFCDSYGAQTSIKTSLTFQSVFAGKTMRNKLLSVLRKCHGLFLDLOVNSLQTVCI 1021	DB	962 LEVFCDSYGAQTSIKTSLTFQSVFAGKTMRNKLLSVLRKCHGLFLDLOVNSLQTVCI 1021
QY	993 NIYKIFLQAFRRFACVQLQPFQDQVRKNTLFTFLGIISQASOCYAILKYNKPMGLKAS 1052	QY	993 NIYKIFLQAFRRFACVQLQPFQDQVRKNTLFTFLGIISQASOCYAILKYNKPMGLKAS 1052	QY	993 NIYKIFLQAFRRFACVQLQPFQDQVRKNTLFTFLGIISQASOCYAILKYNKPMGLKAS 1052
DB	1022 NIYKIFLQAFRRFACVQLQPFQDQVRKNTLFTFLGIISQASOCYAILKYNKPMGLKAS 1081	DB	1022 NIYKIFLQAFRRFACVQLQPFQDQVRKNTLFTFLGIISQASOCYAILKYNKPMGLKAS 1081	DB	1022 NIYKIFLQAFRRFACVQLQPFQDQVRKNTLFTFLGIISQASOCYAILKYNKPMGLKAS 1081
QY	1053 GS---FPEPAHWTLCYQAFLLKLAHSTYKCLLGLPRTAOKLCKLPATWTILKAAA 1109	QY	1053 GS---FPEPAHWTLCYQAFLLKLAHSTYKCLLGLPRTAOKLCKLPATWTILKAAA 1109	QY	1053 GS---FPEPAHWTLCYQAFLLKLAHSTYKCLLGLPRTAOKLCKLPATWTILKAAA 1109
DB	1082 GAAGPLPSEAVQWLCHOAFLLKLAHSTYKCLLGLPRTAOKLCKLPATWTILKAAA 1141	DB	1082 GAAGPLPSEAVQWLCHOAFLLKLAHSTYKCLLGLPRTAOKLCKLPATWTILKAAA 1141	DB	1082 GAAGPLPSEAVQWLCHOAFLLKLAHSTYKCLLGLPRTAOKLCKLPATWTILKAAA 1141
QY	1110 DPALSTDFQTLID 1122	QY	1110 DPALSTDFQTLID 1122	QY	1110 DPALSTDFQTLID 1122
DB	1142 NPALSPDFKTLID 1154	DB	1142 NPALSPDFKTLID 1154	DB	1142 NPALSPDFKTLID 1154
RESULT	4	RESULT	4	RESULT	4
W90251	W90251 standard; Protein; 1132 AA.	W90251	W90251 standard; Protein; 1132 AA.	W90251	W90251 standard; Protein; 1132 AA.
AC	W90251: 24-MAY-1999 (first entry)	AC	W90251: 24-MAY-1999 (first entry)	AC	W90251: 24-MAY-1999 (first entry)
DE	Human catalytic telomerase sub-unit protein.	DE	Human catalytic telomerase sub-unit protein.	DE	Human catalytic telomerase sub-unit protein.
KW	Human; catalytic telomerase subunit; therapy; diagnosis; htc; assay;	KW	Human; catalytic telomerase subunit; therapy; diagnosis; htc; assay;	KW	Human; catalytic telomerase subunit; therapy; diagnosis; htc; assay;
KW	modulator; treatment; inhibit; cellular disorder; death; defect; cancer;	KW	modulator; treatment; inhibit; cellular disorder; death; defect; cancer;	KW	modulator; treatment; inhibit; cellular disorder; death; defect; cancer;
KW	ageing; antisense; neoplastic cell; telomerase-related condition;	KW	ageing; antisense; neoplastic cell; telomerase-related condition;	KW	ageing; antisense; neoplastic cell; telomerase-related condition;
KW	tumour cell.	KW	tumour cell.	KW	tumour cell.
OS	Homo sapiens.	OS	Homo sapiens.	OS	Homo sapiens.
PN	W09859040-A2.	PN	W09859040-A2.	PN	W09859040-A2.
PD	30-DEC-1998.	PD	30-DEC-1998.	PD	30-DEC-1998.
PF	09-JUN-1998; E03468.	PF	09-JUN-1998; E03468.	PF	09-JUN-1998; E03468.
PR	14-APR-1998; DE-016496.	PR	14-APR-1998; DE-016496.	PR	14-APR-1998; DE-016496.
PR	20-JUN-1997; DE-026329.	PR	20-JUN-1997; DE-026329.	PR	20-JUN-1997; DE-026329.
PR	26-MAR-1998; DE-013274.	PR	26-MAR-1998; DE-013274.	PR	26-MAR-1998; DE-013274.
PA	(FARB ) BAYER AG.	PA	(FARB ) BAYER AG.	PA	(FARB ) BAYER AG.
PI	Hagen G, Stegmund H, Weichel W, Wick M, Zubov D;	PI	Hagen G, Stegmund H, Weichel W, Wick M, Zubov D;	PI	Hagen G, Stegmund H, Weichel W, Wick M, Zubov D;
DR	N-PSDB: W72117.	DR	N-PSDB: W72117.	DR	N-PSDB: W72117.
PT	New catalytically active subunit of human telomerase - used in the	PT	New catalytically active subunit of human telomerase - used in the	PT	New catalytically active subunit of human telomerase - used in the
PT	modulation of telomerase activity, particularly for treating cancer	PT	modulation of telomerase activity, particularly for treating cancer	PT	modulation of telomerase activity, particularly for treating cancer

Db 703 PPELYFVKVDVTGAYDITPDRLETEVIAIHK-PQNTYCVRRYAVVQAAHGHVRKAFKS 761  
 Qy 753 QVTTSLDLPYMGOKFLHQLQSDASALRNSVVIQSIISMNESSSSLDFFLHFLRHSVVK 812  
 Db 762 HVSTLTDLPYMRQFVAHQET--SPURDAVVIQSSSLNEASSGLDFVLFRECHHAVR 819  
 Qy 813 IGRDCTQCQIGIPGSSLSLTLSCFCGDMENKFLAEVQRDGLLRVDFVLLVTPHLQ 872  
 Db 820 IRGKSYVOCQIGIPGSSLSLTLSCFCGDMENKFLAEVQRDGLLRVDFVLLVTPHL 879  
 Qy 873 AKTFLTLVHGVPYGYGMINLQKTVNFPVPEGTLGGAAPQLPAHCLFPWCGLLLDTQT 932  
 Db 880 AKTFLRTLVRGVPYGYGMINLQKTVNFPVPEDEALGTAQVMPAHGLFPWCGLLLDT 939  
 Qy 933 LEVCDISGTAQTSIKTSLTFSQVFKAGTKMNRKLLSVLRKLGHLDFLDQVNSLQVCI 992  
 Db 940 LEVQSDYSYARTSIRASLTENRGEKAGNRNRKLFQVLRKCHSLFLDLQVNSLQVCI 999  
 Qy 993 NIYKIFLQAYRHACVQLPQDORVRKNTLTFELGISSQASCCYAILKVNPGMTLKAS 1052  
 Db 1000 NIYKILLQAYRHACVQLPQDORVRKNTLTFELGISSQASCCYAILKVNPGMTLKAS 1059  
 Qy 1053 GS---FPPEAAHWCYQAFLLKLAHSAVYKCLLGLPLRTAQKLLCRKLPEATWTLKAAA 1109  
 Db 1060 GAAGPLPSEAVQWLCHQAFELKLTTRHRTVTVPLLSLRTAQTLQSLPDTTTLTAA 1119  
 Qy 1110 DPALSTDFQITLD 1122  
 Db 1120 NPALPSDFKTLTD 1132

RESULT 5  
 W47000  
 ID W47000 standard; Protein; 1285 AA.  
 AC W47000;  
 DT 13-AUG-1998 (first entry)  
 DE His tagged thiorodoxin moiety and full length hTERT fusion protein.  
 KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;  
 OS Synthetic; cell proliferation; cancer; ageing; ribonucleoprotein.  
 OS Homo sapiens.  
 FH Key  
 FT Location/Qualifiers  
 FT Misc\_difference 119..120  
 FT /note= "enterokinase cleavage site"  
 FT 120..1285  
 FT /label= hTERT  
 FT /note= "full length human telomerase reverse  
 transcriptase"  
 PN G82317891-A.  
 PD 08-APR-1998.  
 PF 01-OCT-1997; 020890.  
 PR 14-AUG-1997; US-915503.  
 PR 01-OCT-1996; US-724643.  
 PR 18-APR-1997; US-844419.  
 PR 25-APR-1997; US-846017.  
 PR 06-MAY-1997; US-851843.  
 PR 09-MAY-1997; US-854050.  
 PR 14-AUG-1997; US-911312.  
 PR 14-AUG-1997; US-912951.  
 PA (GERO-) GERON CORP.  
 PA (UYTE-) UNIV TECHNOLOGY CORP.  
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,  
 PI Morin GB, Nakamura T, Harley CB;  
 DR WPI; 98-171633/16.  
 PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
 PT variants - are useful in the diagnosis, prognosis and treatment of  
 PT cell proliferation conditions especially cancer and ageing  
 PS Example 6; Page 223; 387pp; English.  
 CC The present sequence represents a fusion protein from an example  
 CC of the present invention which describes human telomerase reverse  
 CC transcriptase (hTERT). The present invention also describes the  
 CC following methods: (A) determining whether a test compound is

CC a modulator of hTERT, by detecting the change in hTERT recombinant  
 CC protein or polynucleotide, on administration of the compound;  
 CC (b) preparation of recombinant telomerase by contacting a protein  
 CC preparation of hTERT with a telomerase RNA component; (C) detection  
 CC of the hTERT RNA or protein in a sample by binding a relevant  
 CC probe to the sample and detecting the complex formed or in the case of  
 CC RNA detection, amplifying the product and correlating the presence of  
 CC complex or amplification product with presence of hTERT in the sample;  
 CC and (D) increasing the proliferation of a vertebrate cell by increasing  
 CC hTERT expression; and (E) the use of an agent that causes an increase in  
 CC cell vertebrate cell proliferation to create a medicament that inhibits  
 CC ageing. A protein preparation of hTERT and the polynucleotide encoding  
 CC hTERT can be used in the manufacture of medicaments for inhibiting the  
 CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
 CC used to treat conditions that are associated with high telomerase  
 CC activity. A protein preparation of hTERT can also be used in the new  
 CC methods.  
 CC Sequence 1285 AA;

Query Match 59.3%; Score 3499; DB 1; Length 1285;  
 Best Local Similarity 62.3%; Pred. No. 0;  
 Matches 718; Conservative 122; Mismatches 261; Indels 52; Gaps 13;  
 QY 1 MTRAPRCPAVRSLRLSRVREVMPLATFVRRLLGPEGRRLVQPGDKIYRTLVAAQCLVCMHW 60  
 Db 154 MPRAPRCRAVRSLRLSRVREVMPLATFVRRLLGPEGRRLVQPGDKIYRTLVAAQCLVCMHW 213  
 QY 61 GSQPPADLSFHQVSSSLKELVARVQVRLCERNERNVLAFFGELLNEARGPPMAFTSSVR 120  
 Db 214 DARPPAAPSPFRQVSLKELVARVQVRLCERNERNVLAFFGELLNEARGPPMAFTSSVR 273  
 QY 121 SYLPTNVTIETLRVSGAWMLLSRVGDDLLVLLAHALYLLVPPSCAYQVCGSPLYQICA 180  
 Db 274 SYLPTNVTIETLRVSGAWMLLSRVGDDLLVLLAHALYLLVPPSCAYQVCGSPLYQICA 333  
 QY 181 TTDIWPVSASVPRPRVGRNFTNRLFLQIKSSSRQEPAPKPLPSRGTRKHLSTSTS 240  
 Db 334 ATQARPPPHAS-GPRRLG-----CERAWNSVREAGVPLGLPAPGARRGGSASRS 384  
 QY 241 VPSAKKARCPVPRVEEGP-----HRQVLPSPGKSW-VPSPARSDEVTAKDLSSK 292  
 Db 385 LPLPRPRGAPEPPTPVGGGWAHPGTRGSPDGRFCVVSAPAR-----PABEATSLE 439  
 QY 293 GKVSLSLS-GSVCKKHKPSSTSLSPPRONAFOLRP-FIETRHFLYSGDQERLNPSPF 350  
 Db 440 GALSOTRSHSPVGRQHHAGPPTSRRPDPWDTPCPVYAEIKHFLYSSGD-KEQLRPSF 498  
 QY 351 LLSNLPNLTGARLVEIIFLGSRRPTSGPLCRTHRSRRYWMRPLFQQLLVNHAECQY 410  
 Db 499 LLSLRPLSTGARLVEIIFLGSRRPTSGPLCRTHRSRRYWMRPLFQQLLVNHAECQY 558  
 QY 411 VRLLSHCRFTANQOVTDAL-----NTSPHMLDLRLHSSPHQVY 452  
 Db 559 GVLLKTHCPLEAA---VTPAAGVCAREKPGSVAAPEEDDPRLVQLLRHSSPHQVY 615  
 QY 453 GFLRACLVVYASLWGTNRHNERFFKNLKKFISLGRYKLSLQELMKWKKVDECHWLSR 512  
 Db 616 GFVRACLRLVPPGLWGSRRHNERFFKNLKKFISLGRYKLSLQELMKWKKVDECHWLSR 675  
 QY 513 SPGKDRYPAAEHLRLERILATFLEWMDTVVQLLSFFYITESTFQKNLFFYRKSVWS 572  
 Db 676 SPGVGCVPAAEHLRLERILATFLEWMDTVVQLLSFFYITESTFQKNLFFYRKSVWS 735  
 QY 573 KLSQIGVQHLRLERILATFLEWMDTVVQLLSFFYITESTFQKNLFFYRKSVWS 632  
 Db 736 KLSQIGVQHLRLERILATFLEWMDTVVQLLSFFYITESTFQKNLFFYRKSVWS 795  
 QY 633 RALGRRKQAQHTQRLKTLFSLMNYERTKPHLMGSSVGLGMDIYRTWRAFLVRALDQ 692  
 Db 796 RTFRREKRAERLTSRVKALFSVLNYSERARRPGLLGASVLGLDDDIHRAWRTFVLVRAQDP 855  
 QY 693 TPRMYFVKADVTGAYDAIPQGLVEVVANNIRHSESTYICIRQYAVVRRDSQGVHKSFR 752

Db 856 PPELYFKVDVTGAYDITIPQDRTEVIASLIK-PONTYCVRRYAVVQKAAGHVRKAFKS 914  
QY 753 QVTTLSLQPMQOFLKHLQSDASALRNSVVEIQSISMNESSSDFDFLHFLRHSVVK 812  
Db 915 HVSTLTDLPQYMQFVAHLQET--SPLRDVAVIEQSSSLNEASSGLDFVLRPMCHHAVR 972  
QY 813 IGDRCYTCCOGIPQGSLSLLCSLFCGDMENKLFQVQSDGGLLRFVDDFLAVTPHLQ 872  
Db 973 IRKSYVQCOGIPQGSLSLLCSLFCGDMENKLFQVQSDGGLLRFVDDFLAVTPHLH 1032  
QY 873 AKTFLSLVHGVPYEGCMINLQKTVNFPVPGTGLGAAPYQLPAHGLFPCWGLLDTQT 932  
Db 1033 AKTFLRLVGVPEYGCVNLRKTVNFPVDEALGTAFAVQMPAHGLFPCWGLLDTRT 1092  
QY 933 LEVFCDSYGAQTSIKTSLTFQSVFKAGTKMRNKLKSVLRKCHGLFLDLQVNSLQTVCI 992  
Db 1093 LEVQSDYSYARTSIRASLTFRNGFKAGRNMRKLFGLVLRKCHSLFLDLQVNSLQTVCT 1152  
QY 993 NIYKIFLQAYRHFACVIOLPEDQVRKNTFFLGIITSSOASCCYAILKVNPGMTLKAS 1052  
Db 1153 NIYKILLQAYRHFACVIOLPEDQVRKNTFFLGIITSSOASCCYAILKVNPGMTLKAS 1212  
QY 1053 GS---FPPEAAHLCYQAFLLKLAHVIYKCLGLPRTAQKLCRLKLPKATMTILKAAA 1109  
Db 1213 GAAGPLPSEAVOWLCHQAFLLKLRHRTVYVPLLSLRTAQTLQSLRKLPGTTLTALEAAA 1272  
QY 1110 DPALSTDFQTILD 1122  
Db 1273 NPALPSDFKTILD 1285

RESULT 6  
W71376  
ID W71376 standard; Protein; 1132 AA.  
AC W71376;  
DE Human telomerase catalytic subunit referred to as hstr2.  
KW Catalytic subunit; human; telomerase; telomere maintenance;  
diagnosis; treatment; cancer.  
OS Homo sapiens.  
PN W09837181-A2.  
PD 27-AUG-1998.  
PF 20-FEB-1998; U03404.  
PR 30-OCT-1997; US-064322.  
PR 20-FEB-1997; US-038750.  
PR 20-MAY-1997; US-047151.  
PR 01-AUG-1997; US-054549.  
PR 14-AUG-1997; US-055762.  
PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.  
PI Counter CM, Meyerson M, Weinberg RA;  
DI WPI: 98-495367/42.  
DR N-PSDB: V60320.  
PT New isolated human telomerase catalytic sub-unit gene - used to  
develop products for increasing or reducing the life span of cells  
such as cancer cells or transformed cells  
PS Claim 5; Fig 6; 96pp; English.  
CC The present sequence represents the catalytic subunit of a human  
telomerase holoenzyme. Disruption of the telomerase gene alters  
telomere maintenance. The DNA is essential for telomerase activity,  
and the protein is physically associated with telomerase and a  
constituent of active telomerase complex. The products can be used  
for increasing or reducing the lifespan of cells such as cancer cells  
or transformed cells. They can also be used in the diagnosis and  
treatment of malignancies. In addition, cells with a longer lifespan  
can be transplanted into or grafted onto an individual (e.g. as skin  
grafts, as systems for delivery of therapeutic proteins, such as hormones  
and enzymes), to whom they provide therapeutic benefit.  
SQ Sequence 1132 AA;

Query Match 59.3%; Score 3498; DB 1; Length 1132;  
Best Local Similarity 62.3%; Pred. No. 0;

Matches 718; Conservative 122; Mismatches 261; Indels 52; Gaps 13;  
QY 1 MTRAPRCVAVRSLRSRYREVWPLATFVRRLLGPEGRLVQDQPKIYRTLVQAQCLVCHWH 60  
Db 1 MPRAPRCVAVRSLRSRYREVWPLATFVRRLLGPEGRLVQDQPKIYRTLVQAQCLVCHWH 60  
QY 61 GSQPPADLSFHQVSSSLKELVARVQRCERNERNVLAFFGELLNEARGGPPMAETSSVR 120  
Db 61 DARPPPAAPSQVQVSCLELVARVQRCERNERNVLAFFGELLNEARGGPPMAETSSVR 120  
QY 121 SYLPTNVTIETLRSVGAWMLLSRVGDDLLVLAHALYLLVPPSCAYQVCSPLIYQICA 180  
Db 121 SYLPTNVTIDALRGSGAWMLLSRVGDDLLVLAHALYLLVPPSCAYQVCSPLIYQICA 180  
QY 181 TTDIWPSVSYRTPRPVGRNFTNLRFLLQIKSSRSQEQAPKPLALPSRGTKRHLSTSTS 240  
Db 181 ATQARPPPHAS-GPRRLTG-----CERAWNHSVREAGVPLGLPAPGARRRGGSASRS 231  
QY 241 VPSAKKARCYVPVVEEGP-----HROVLTPTSGKSW-VSPARSPEVPTAEKDLSSK 292  
Db 232 LPLPKPRRGAAPERTVPGQGSWAHPGTRGSDRGECVVSAPAR-----PAEATSLSE 286  
QY 293 GKYSDDLSS-GSVCCCKHKPSSTLSLSPPRQNAFQLRP-FIETRHFLYSRGDQERLNPFS 350  
Db 287 GALSGRHSHPSVGRHQHAGPPSTSRPPPDWTPCPVVAETKHFLYSSGD-KEQLRPSF 345  
QY 351 LLSNLQPNLTGARRUVEIIFLGSRPRTSGPLCTHRLSRRYQWRPFLFOQLLVNHAECY 410  
Db 346 LLSLSRSLTGARRLVETIFLGSRPWMPGTPRRLPLPQRYQWRPFLFELIGNHAQCPY 405  
QY 411 VRLLSHCRFTANQOVTDAL-----NTSPHMLDMLLRHSSPWQY 452  
Db 406 GVLLTKHCPURAA---VTPAAGVCAREKPGQSVAAPEEEDTPRRLVQLLRQHSSPWQY 462  
QY 453 GFLRACLCYVVSASLGMTRHNRFFKNLKFISLQYKLSLQELQELMWMKVDCHWLRS 512  
Db 463 GFVRACLRLRVPPLGMSRHNRRFLRNTKFKISLQKHAHLSLQELTWKMSVRCGLAWLR 522  
QY 513 SPQKDRVPAAEHLRERILATFLFWMIDTVYVOLLRSFFVITESTFQKRLFFYRSVWS 572  
Db 523 SPGVGCVPAAEHLRERILAKFLHMLMSVYVVELLSFFVYTTTTFQKRLFFYRSVWS 582  
QY 573 KLOSIGVROHLERVLRELSEVRRHQDTWLAIPCRLRFIPKPNGLRP IVNNYSMGT 632  
Db 583 KLASIGIRQHLKRVQLRELSEAEVRQREARPAALLTSRLRFIPKPNGLRP IVNNYVYVGA 642  
QY 633 RALGRRKQAOHFTORLKTFLSMLNRYERTKPHLMGMSVLGMNDIYRTWRAFLVRALDQ 692  
Db 643 RTFRERKRAERLTSRVKALFVNLNERARRPGLIGASVLGLDDIHRARWTFVLRVAQDP 702  
QY 693 TPRMYFVKADVTGAYDAIPOGKLVVEVANNIRHSESTYCIROYAVVRRDSQGVHKSFR 752  
Db 703 PPELYFKVDVTGAYDITIPQDRTEVIASLIK-PONTYCVRRYAVVQKAAGHVRKAFKS 761  
QY 753 QVTTLSLQPMQOFLKHLQSDASALRNSVVEIQSISMNESSSDFDFLHFLRHSVVK 812  
Db 762 HVSTLTDLPQYMQFVAHLQET--SPLRDVAVIEQSSSLNEASSGLDFVLRPMCHHAVR 819  
QY 813 IGDRCYTCCOGIPQGSLSLLCSLFCGDMENKLFQVQSDGGLLRFVDDFLAVTPHLQ 872  
Db 820 IRKSYVQCOGIPQGSLSLLCSLFCGDMENKLFQVQSDGGLLRFVDDFLAVTPHLH 879  
QY 873 AKTFLSLVHGVPYEGCMINLQKTVNFPVPGTGLGAAPYQLPAHGLFPCWGLLDTQT 932  
Db 880 AKTFLRLVGVPEYGCVNLRKTVNFPVDEALGTAFAVQMPAHGLFPCWGLLDTRT 939  
QY 933 LEVFCDSYGAQTSIKTSLTFQSVFKAGTKMRNKLKSVLRKCHGLFLDLQVNSLQTVCI 992  
Db 940 LEVQSDYSYARTSIRASLTFRNGFKAGRNMRKLFGLVLRKCHSLFLDLQVNSLQTVCT 999  
QY 993 NIYKIFLQAYRHFACVIOLPEDQVRKNTFFLGIITSSOASCCYAILKVNPGMTLKAS 1052  
Db 1000 NIYKILLQAYRHFACVIOLPEDQVRKNTFFLGIITSSOASCCYAILKVNPGMTLKAS 1059







ID W56101 standard; Protein; 1405 AA.  
AC W56101;  
DT 13-AUG-1998 (first entry)  
DE Enhanced green fluorescent protein and hTERT fusion protein.  
KW Human; telomerase reverse transcriptase; hTERT; PRT; diagnosis;  
KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein.  
OS Synthetic.  
OS Homo sapiens.  
FH Key  
FT Location/Qualifiers  
FT 1..250  
FT /note= "enhanced green fluorescent protein fragment"  
FT 276..1405  
FT /note= "hTERT protein fragment"  
PN GB2317891-A.  
PD 08-APR-1998.  
PF 01-OCT-1997; 020890.  
PR 14-AUG-1997; US-915503.  
PK 01-OCT-1996; US-724643.  
PK 18-APR-1997; US-844419.  
PR 23-APR-1997; US-846017.  
PR 06-MAY-1997; US-851843.  
PR 09-MAY-1997; US-854050.  
PR 14-AUG-1997; US-911312.  
PR 14-AUG-1997; US-912951.  
PA (GERO-) GERON CORP.  
PA (UYTE-) UNIV TECHNOLOGY CORP.  
PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,  
PI Morin GB, Nakamura T, Harley CB;  
DR WPI: 98-171633/16.  
PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
PT variants - are useful in the diagnosis, prognosis and treatment of  
PT cell proliferation conditions especially cancer and ageing  
PS Example 15: Page 269-270; 387pp; English.  
CC The present sequence represents a fusion protein from an example  
CC of the present invention which describes human telomerase reverse  
CC transcriptase (hTERT). The present invention also describes the  
CC following methods: (A) determining whether a test compound is  
CC a modulator of hTERT, by detecting the change in hTERT recombinant  
CC protein or polynucleotide, on administration of the compound;  
CC (B) preparation of recombinant telomerase by contacting a protein  
CC of the hTERT RNA or protein in a sample by binding a relevant  
CC RNA detection, amplifying the product and correlating the presence of  
CC and (D) increasing the proliferation of a vertebrate cell by increasing  
CC hTERT expression; and (E) the use of an agent that causes an increase in  
CC cell vertebrate cell proliferation to create a medicament that inhibits  
CC ageing. A protein preparation of hTERT and the polynucleotide encoding  
CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
CC used to treat conditions that are associated with high telomerase  
CC activity. A protein preparation of hTERT can also be used in the new  
CC methods.  
SQ Sequence 1405 AA;

Query Match 58.7%; Score 3463; DB 1; Length 1405;  
Best Local Similarity 61.9%; Pred. No. 0;  
Matches 713; Conservative 121; Mismatches 266; Indels 52; Gaps 13;  
QY 1 MTRAPCAPVRLSLRSRYEWPLATFVRLGPGGRRLVQDPKPIRLVLAQCLVCMHW 60  
DB 276 MPRAPCRVRLSLRSRYREVPLATFVRLGPGGRRLVQDPKPIRLVLAQCLVCMHW 60  
QY 61 GSOPPPADLSFHOVSSLSKELVARVORLQERNERNYLAFCFELLNEARGPPMAFTSSVR 120  
DB 336 DARPAPPAFQVSCIKELVARVORLQERNERNYLAFCFELLNEARGPPMAFTSSVR 120  
QY 121 SYLPTVETLRVSGAWMLLSRVGDDLLVLLAHGALVLLPPSCAYQVCGSPLYQICA 180  
DB 396 SYLPTVETLRVSGAWMLLSRVGDDLLVLLAHGALVLLPPSCAYQVCGSPLYQICA 180

QY 181 TTDIWPVSYSASYRTPRPVGRNFTNLRLQKSSROBAPKPLALPSRGTKRHLSTSTS 240  
DB 456 ATOARPPPHAS-GPRRLG-----CERAWNHVSREAGVPLGAPGARRGGSSASRS 506  
QY 241 VPSAKKARCYPVPRVEEGP-----HRQVLPTPSGSKM-VPSPARSPVEPTAKDLSSK 292  
DB 507 LPLPKRPRGAAPPERTPVGGQSWAHPGRTRGSDRGFCVSPAR-----PAEATSLE 561  
QY 293 GKVSLSLS-GSVCKKHKPSSTLSLSPRONAFOLRFETRIHFLSLYSRGQERLPSFL 351  
DB 562 GALSCTRHSHSVGRQHAGPPSTSRPRWDTPCPVYAEATKFLYSSGD-KEQLRPSFL 620  
QY 352 LSNLQPNLTGARRLVEILFGLSRPRTSGPLCRTHLSRRYQWQMRPLFQOQLVNHAEQVY 411  
DB 621 LSSLRPSLTGARRLVETIFLGRPMMEGTPLRLPLPQRYQWQMRPLFLELLGNHACPYG 680  
QY 412 RLLRSHCFRTANQVTDAL-----NTSPHMLDLRLHSSSPQVYVG 453  
DB 681 VLLKTHCPLRAA---VTPAGVCAREKPGQSVAAPEEDTDPRLVOLLROHSSPQVYVG 737  
QY 454 FLRACICKVYSASLWGTNRHNERFFKNLKFISLGRYKGLSLQELMKMKVEDCHWLRS 513  
DB 738 FVRACLRRLVPPGLWGRHNERFLRNTKFIISLGNHAKLSLOELTKWMSVRDCAWLRS 797  
QY 514 PGKDRVPAAEHRLRLRLATLFTLWMDTVVQLLSFFYITESTFOKNLFFYKRSVMSK 573  
DB 798 PGVGCVPAAEHLREILAKFLHLMMSVYVVELLSFFYITESTFOKNLFFYKRSVMSK 857  
QY 574 LQSIGVQHLEVRVRLRELSQEEVRRHODTWLAMPICRLRFIPKNGLRPIVNMVSMGTR 633  
DB 858 LQSIGVQHLEVRVRLRELSQEEVRRHODTWLAMPICRLRFIPKNGLRPIVNMVSMGTR 633  
QY 634 ALGRRKQAOHFTORLKTLESMNLNYERTKHPHMGSSVGMNDIYRTWRAFLVRALDQT 693  
DB 918 TFRERARERLTSRVKALEFSLNYERARPPGLGASVLGLDDIHRWARTFVLVRADQP 977  
QY 694 PRMYFVKDVTGAYDAIPQGLVEVANNIRHSESTYCIROYAVVVRDROGQVHKSPFRQ 753  
DB 978 PELYFVKDVTGAYDIIPDRLETVIASIK-PONTYCVRIYVQKAAHGVKAKFSH 1036  
QY 754 VTTLSLQPMYQOFLKHLQSDASALRNSVIEOSISMNESSSLSLDFFLHFLRHSVVKI 813  
DB 1037 VSTLDLQPMYQOFLKHLQSDASALRNSVIEOSISMNESSSLSLDFFLHFLRHSVVKI 813  
QY 814 GRCYTCQCGIPQGSSTLCLSCFCGDMENKLAFAVORDGLLLRFVDDFLVTPHLDQA 873  
DB 1095 RGSYVQCGIPQGSSTLCLSCFCGDMENKLAFAVORDGLLLRFVDDFLVTPHLDQA 873  
QY 874 KTFSLTVHGVPEYGCMLNOKTVVVPFPGTLGGAPYQLPARCLFPWCGLLDFTOL 933  
DB 1155 KTFSLTVHGVPEYGCMLNOKTVVVPFPGTLGGAPYQLPARCLFPWCGLLDFTOL 933  
QY 934 EYFCDSYGAOTSITKTSFQSVFKAGTKRNLKLSVLRLKCHGLFLDLQVNSLQTVGIN 993  
DB 1215 EVQSDYSYARTSRASVTFNRGFGAKNRMRKLFGLVRLKCHSLFLDLQVNSLQTVGIN 1274  
QY 994 IYKFLQAYRFHACVITQLPDQVRKRLNLTFFLGIISQASCCVATILKYNKPGMTLKASG 1053  
DB 1275 IYKFLQAYRFHACVITQLPDQVRKRLNLTFFLGIISQASCCVATILKYNKPGMTLKASG 1053  
QY 1054 S---FPPEAAHLCYQAFLLKLAHSHVYIKCLLGPRTAQKLCRKLPEATMTILKAAAD 1110  
DB 1335 AAGPLPSEAVQWCHQAFLLKLRHRTVYVPLGLSL-TAQTLRSRKLPGTTLTALEAAN 1393  
QY 1111 PALSTDFQTILD 1122  
DB 1394 PALPSDFKTILD 1405  
RESULT 10  
W47006  
ID W47006 standard; Protein; 1150 AA.  
AC W47006;



Mon Aug 16 10:21:26 1999

13-AUG-1998 (first entry)  
 DE Glutathione-S-transferase and hTERT fusion protein 6.  
 KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;  
 KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein.  
 OS Synthetic.  
 OS Homo sapiens.  
 PN GB2317891-A.  
 PD 08-APR-1998.  
 PF 01-OCT-1997; 020890.  
 PR 14-AUG-1997; US-915503.  
 PR 01-OCT-1996; US-724643.  
 PR 18-APR-1997; US-844419.  
 PR 25-APR-1997; US-846017.  
 PR 06-MAY-1997; US-851843.  
 PR 09-MAY-1997; US-854050.  
 PR 14-AUG-1997; US-911312.  
 PR 14-AUG-1997; US-912951.  
 PA (GERO-) GERON CORP.  
 PA (UYTE-) UNIV TECHNOLOGY CORP.  
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,  
 PI Morin GB, Nakamura T, Harley CB;  
 DR WPI; 98-171633/16.  
 PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
 PT variants - are useful in the diagnosis, prognosis and treatment of  
 PT cell proliferation conditions especially cancer and ageing  
 CC Example 6; Page 231-232; 387pp; English.  
 CC The present sequence represents a fusion protein from an example  
 CC of the present invention which describes human telomerase reverse  
 CC transcriptase (hTERT). The present invention also describes the  
 CC following methods: (A) determining whether a test compound is  
 CC a modulator of hTERT, by detecting the change in hTERT recombinant  
 CC protein or polynucleotide, on administration of the compound;  
 CC (B) preparation of recombinant telomerase by contacting a protein  
 CC preparation of hTERT with a telomerase RNA component; (C) detection  
 CC of the hTERT RNA or protein in a sample by binding a relevant  
 CC probe to the sample and detecting the complex formed or in the case of  
 CC RNA detection, amplifying the product with presence of hTERT in the sample;  
 CC complex or amplification product with presence of hTERT in the sample;  
 CC and (D) increasing the proliferation of a vertebrate cell by increase in  
 CC hTERT expression; and (E) the use of an agent that causes an increase in  
 CC cell vertebrate cell proliferation to create a medicament that inhibits  
 CC ageing. A protein preparation of hTERT and the polynucleotide encoding  
 CC hTERT can be used in the manufacture of medicaments for inhibiting the  
 CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
 CC used to treat conditions that are associated with high telomerase  
 CC activity. A protein preparation of hTERT can also be used in the new  
 CC methods.  
 SQ Sequence 1150 AA;  
 Query Match 56.3%; Score 3320; DB 1; Length 1150;  
 Best Local Similarity 61.2%; Pred. No. 7.8e-299;  
 Matches 708; Conservative 122; Mismatches 263; Indels 64; Gaps 23;  
 QY 1 MTRAPCPAVRSLLSRVREWPPLATFVRRLGPEGRRLLVQGDPRKIYRTLVQAQCLVCMHW 60  
 DB 1 MPRAPPCRAVRSLLSRVREWPPLATFVRRLGPEGRRLLVQGDPRPAFALVQAQCLVCPW 60  
 QY 61 GSOPPPADLSFHOVSKKELVARVQRLCERNERNVLAFFGELLNEARGGPPMAFTSSVR 120  
 DB 61 DARPPAAPSFQVSCLEKELVARVQLRQRCAGKKNVLAFFGALLDARGGPPMAFTSSVR 120  
 QY 121 SYLPTNTVIETLVSGAWMLLSRVGDDLLVYLALHCAKLVLLVPPSCAYQVCGSPLYQICA 180  
 DB 121 SYLPTNTVIETLVSGAWMLLSRVGDDLLVYLALHCAKLVLLVPPSCAYQVCGSPLYQICA 180  
 QY 181 TTDWPSVSAVSRPTVPVGRNFTNLRFLOQIKSSSRQEAQKPLALPSRGTKKHLSLTSTS 240  
 DB 181 ATQARPPPHAS-GPRRLG-----CERAWNHSVREAGVPLGLPAPGARRRGGSASRS 231  
 QY 241 VPSAKKARCVPRVEBGP-----HRQVLPPTPSGKSW-VPSPARSPVPTAEKDLKSL 292  
 DB 232 LPLPKRPRRGAAPERTPTVGGGSAWHPGTRGFSRGFCVSPAR-----PAEEATSL 286

293 GKVSLSLS-GSVCKCHKPSSTLSLPPRQAFQRLP-FIETRHLYSRGQDQRLNPSF 350  
 287 GALSGRTHSHPSVGRHQHAGPSTSRPPRWDTPCAPPVYATKHFLLSSGD-KEQLRPSF 345  
 351 LLASNQPNLTGARRLVEIIFLGSRRPTSGPICRTHLSRRYWMQRPFLQOQLLVNHAECQY 410  
 346 LLSLRPSLTGARRLVETIFLGS-PWMEFTPRRLPRLPQRYWMQRPFLLELLGNHAQCFY 404  
 411 VLLRSHCFRTANQOVTDAL-----NTSPPHMLDLRLHSSPWQVY 452  
 405 GVLLKTHCPLRAA---VTPAAGVCAREKPGQSVAAPEEEDTPRLVOLLRQHSSPWQVY 461  
 453 GFLRACILCKVVSASL-WGTRHNERFFKMLKPKFISLGKYGKLSQLQELMKMKVECHWLRL 511  
 462 GFVRACLRRLVPPGLVPGLEVSGRHNERRLRNTRKFKFISLGKHAQLSLOELTWKMSVROCAMLR 521  
 512 SSPGKORVPAAEHLRERILATFLWMDTYVYVOLLRSFFYITESTFQKNRLFYFKSYW 571  
 522 RSPGVCVPAAEHLRERILAKELHLMVSYVVELLSRFW-TEITFQKNRLFYFKSYW 580  
 572 SKLQSIGVROHLRERLRELSQEVRRHQDTWLAMPICRLRIFPKPKNGLRPIVNNYSYMG 631  
 581 SKLQSIGVROHLRERLRELSQEVRRHQDTWLAMPICRLRIFPKPKNGLRPIVNNYSYMG 640  
 632 TRALGRKQAHQHTQRLKTLFSLMNYERTKPHLMGSSVLGMNDIYRTWRAFLVLRALD 691  
 641 ARTEREKRAERLTSRYKALFVNLNERARRPGLLGASVLGDDIHRAMRTFVLRVAQD 700  
 692 QTPRMVFKADVTGAYDAIPQGLVEVYVANNIRHSESTYCIQYAVYVRSDSQOVHKSFR 751  
 701 PPPELYFKVDVTGAYDTIPDRLETVIASLIK-PQNTYCVRRYAVYVQKAAHGHVRKAFK 759  
 752 RQVTLSDLPYMGQFLKHLQDSASALRNSVYEQSISM-NESSSLDFDFHFLRHVS 810  
 760 SHVSTLTLPYMRQFVAHQET--SPLRDAVYEQSSSLETSASSGLDFVFLRFMCHHA 817  
 811 VKIGRCVTCQCGIPQGSSTLSCFCFQDMENKLFQVQROGLLRFVDDFLVLTPHL 870  
 818 VERGS--WQCGIPQGSSTLSCFCFQDMENKLFQVQROGLLRFVDDFLVLTPHL 875  
 871 DOAKTFTLTVHGVPEYCGMNLQKTVVNFVEPTGTTGGAAPYOLPAHCLFPW-CGLILD 929  
 876 THAKTFTIRLVG-PEYGVVNLKTV--FPVEDALGGTAFOVMPAHGLEFPWVCGLLD 932  
 930 TOTLEVCDSYATQTSIKTSLTQSFQVFKAG-TMRNKLKLVRLKCHGLFDLQVNSLQ 988  
 933 TRTLEVSQDESSYATSIASLTFNRFKAGRETMRRLKLVRLKCHGLFDLQVNSLQ 992  
 989 TVCINIYKIFLQAYRHACVQLPFDQVRKNLTFLLGISSQASCCYAILKVNPKMT 1048  
 993 TVCTNIYKILLQAYRHACVQLPFDQVRKNLTFLLGISSQASCCYAILKVNPKMT 1052  
 1049 LRAGS---FPPEAAHLCYQAFLLKLAHSHYIYKLLGPLRTAOKLCKRLKPEATMTIL 1105  
 1053 LGKAGAGPLPSEAVOWLCHQAFLLKLTHTRYTYVPLLGSLRTAQTQLSRKLPGTTLTAL 1112  
 1106 KAAADPALSTDFQITLD 1122  
 1113 EAAANPALS-DFKTILD 1128

RESULT 11  
 W61349 standard; Protein; 949 AA.  
 ID W61349;  
 AC W61349;  
 DT 12-OCT-1998 (first entry)  
 DE Human telomerase protein 2 (TP2) partial polypeptide.  
 KW TP2; human; telomerase protein 2; cancer; AIDS; ageing; therapy.  
 OS Homo sapiens.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Protein 640..940  
 FT /note= "Claim 22"

FT Protein 1. .563  
 FT /note= "Claim 24"  
 FT Protein 1. .640  
 FT /note= "Claim 24"  
 FT Protein 696. .940  
 FT /note= "Claim 24"  
 FT Protein 696. .953  
 FT /note= "Claim 24"  
 FT Region 582. .587  
 FT Region 644. .648  
 FT Region 852. .857  
 FT Region 884. .8894  
 PN W09821343-A1.  
 PD 22-MAY-1998.  
 PF 13-NOV-1997; U21248.  
 PR 16-OCT-1997; US-951733.  
 PR 15-NOV-1996; US-871189.  
 PR 11-JUN-1997; US-873039.  
 PA (AMGE-) AMGEN CANADA INC.  
 PI HARRINGTON LA, ROBINSON MO;  
 DR WPI: 98-297946/26.  
 DR N-PSDB: V27872.  
 PT New nucleic acid encoding human telomerase protein-2 - used for  
 PT regulating telomerase activity, e.g. for treating cancer or acquired  
 PT immune deficiency syndrome  
 PS Claim 1d; Fig 6; 150pp; English.  
 CC This polypeptide comprises a large portion of human telomerase  
 CC protein 2 (TP2), a novel protein of the telomerase complex. Its  
 CC amino acid sequence was deduced from partial cDNA clone 32 (see  
 CC V27872), obtained from a human colon tumour cell line LIM1863 cDNA.  
 CC A full-length polypeptide sequence (see W61350) is also disclosed.  
 CC Expressing TP2 in a cell is used to increase telomerase activity and  
 CC thus proliferation for treatment of e.g. HIV infection, AIDS and  
 CC ageing disorders, while expressing an inactive mutant of TP2 (or  
 CC molecule antisense to the gene) is used to decrease telomerase  
 CC activity, e.g. for treatment of cancer. TP2 polypeptides can also  
 CC be used to screen for agents that inhibit TP2 activity or its  
 CC binding to TRIP1 (see W61347) or telomerase RNA, potentially useful  
 CC therapeutically, also to raise specific antibodies useful in  
 CC immunosays and therapeutically as inhibitors. Also contemplated  
 CC are transgenic animals in which the TP2 gene has been inactivated or  
 CC is overexpressed. TP2 polypeptides are administered i.v., s.c. or  
 CC orally, or they are delivered from engineered cells or gene therapy  
 CC vectors.  
 SQ Sequence 949 AA:

Query Match 47.38; Score 2790.5; DB 1; Length 949;  
 Best Local Similarity 60.88; Pred. No. 7.7e-250;  
 Matches 576; Conservative 106; Mismatches 217; Indels 49; Gaps 12;

QY 1 MTRAPRCVRSLLRSRYREYVPLATFVRRLLGPGRRVLPQDPIYRTLVQAQCLVCMHW 60  
 DB 23 MPAPRCVRSLLRSRYREYVPLATFVRRLLGPGRRVLPQDPIYRTLVQAQCLVCMHW 82  
 QY 61 GSQPPADLSHQVSSLSKELVARVQVRLCERNERNVLAFFGELLNEARGGPPMAFTSSVR 120  
 DB 83 DARPPPAAPSPFQVSLKELVARVQVRLCERNERNVLAFFGELLNEARGGPPMAFTSSVR 142  
 QY 121 SYLPTNTVETLRVSGAMWLLSRVGGDLLVYLLAHCAALYLLVPSCAYQVCGSPLOYICA 180  
 DB 143 SYLPTNTVETLRVSGAMWLLSRVGGDLLVYLLAHCAALYLLVPSCAYQVCGSPLOYICA 202  
 QY 181 TTDIWPVSASVRYRTPVGRNFTNRLFLQIKSSRSROEAPKPLAPSRGKRLHLSLTS 240  
 DB 203 ATQARPPPHAS-GPRRLG-----CERANNHSVREAGVPLGLPAPARRGGGASRS 253  
 QY 241 VPSAKKARCVPRVREGP-----HRQVLPVPSGSKW-VPSPARSPVEPTAEKDLSSK 292  
 DB 254 LPLKPRRGAAPERTPVGQSWAHPPGRTRGSDRGFCWSPAR-----PAEATSLE 308  
 QY 293 GKVSLSLS-GSVCKKHKPSSTLLSPRQNAQLRP-FIETRHFLYSGDGOERLNPSE 350

Db 309 GALSCTRHSHPSVGRQHAGPSTSRPPWDTPCPPVYAEKHFLYSSGD-KEQLRSEF 367  
 QY 351 LLSNLPNLTGARRLVEIFLGSRRPTSGPLCRHRLSRRYQWMPPLFQQLLVNHAEOY 410  
 Db 368 LLSLRPLSLTGARRLVETIFLGSRRMPGTPRRLPRLPQRYQWMPPLFLELLGNAOCPY 427  
 QY 411 VRLLSHCRFTANOQVTDAL-----NTSPHMLDMLRLHSSPMOYV 452  
 Db 428 GVLLKTHCPLEAA---VTPAAGVCAREKPGQSVAAPEEEDTPRRLVQLLRQHSPPQYV 484  
 QY 453 GFLRACLCCKVVSASLWGTNRHNRFFKFKFISLGYKLSLQELMKWKKVEDECHWLS 512  
 Db 485 GFVRACLRLRVPPLGWSRHRERFLNTKFKFISLGHAKLSLQELTWKWSVRCAWLR 544  
 QY 513 SPGRDVPAAEHRRLERILATFLWMDTYVYVQLRSEFFYITESTFQKNLFFYKRSVWS 572  
 Db 545 SPGVGCVPAAEHRLREELAKFLHLSVYVVELLSFFYVTTFTQKNLFFYKRSVWS 604  
 QY 573 KLSIGVROHLERVLRELSEOEYVRHODTWLAMPICRLRFIPKPNGLRIVNMSYNGT 632  
 Db 605 KLSIGIFQHLKRVQLRELSEAEVRQREARPPALLTSRLRFIPKPNGLRIVNMSYNGT 664  
 QY 633 RALGRRQAOHTORLKTFLSMNYERTKPHLMGSSVLGMNDIYRTWRAFLVRALDQ 692  
 Db 665 RTFRREKRAERLTSRVKALFSVLNRYERARRPGLLGASVLGLDDIHRARWTFVLVRAQDP 724  
 QY 693 TPRMYFVKADVGTAYDAIPOGKLVVVVANNIRHSESTYCIROYAVVRRDSQGVHKSFR 752  
 Db 725 PPEIYFVKVDVGTAYDTIPQDRLTEVIAIIRK-PONTYCVRRYAVVQKAHGHVRAKFS 783  
 QY 753 QVTLSLQPMYQFQKHLQSDASALRNWSVTEQSTSMNESSSLSLDFELHFLRHSVVK 812  
 Db 784 HVSTLTDLQPMYQFVAHLQET--SPLRDVAVTEQSSLSNEASSGLDFVLFREWFCHHAVR 841  
 QY 813 IGRCYTCOCGIPQGSLSLTLSCFCGDMENKFAEYORGLLRLFRVDDFLVTPHLQ 872  
 Db 842 IRGKSVYQCGIPQGSLSLTLSCFCGDMENKFAEYORGLLRLFRVDDFLVTPHLQ 901  
 QY 873 AKTFLSLVHGVEYGCMLNOKTVNFEVPGTLCGAAPYQLPAHCL 920  
 Db 902 AKTFLSLVHGVEYGCMLNOKTVNFEVPGTLCGAAPYQLPAHCL 949

RESULT 12  
 W46997  
 ID W46997 standard; Protein: 807 AA.  
 AC W46997;  
 DT 13-AUG-1998 (first entry)  
 DE Human telomerase reverse transcriptase Deltal82 variant.  
 KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;  
 OS Synthetic.  
 OS Homo sapiens.  
 PN GB2317891-A.  
 PD 08-APR-1998.  
 PF 01-OCT-1997; 020890.  
 PR 14-AUG-1997; US-915503.  
 PR 01-OCT-1996; US-724643.  
 PR 18-APR-1997; US-844419.  
 PR 25-APR-1997; US-846017.  
 PR 06-MAY-1997; US-851843.  
 PR 09-MAY-1997; US-854050.  
 PR 14-AUG-1997; US-911312.  
 PR 14-AUG-1997; US-912951.  
 PA (GERO-) GERON CORP.  
 PA (UYTE-) UNIV TECHNOLOGY CORP.  
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,  
 PI Morin GB, Nakamura T, Harley CB;  
 DR WPI: 98-171633/16.  
 DR N-PSDB: V22382.  
 PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
 PT variants - are useful in the diagnosis, prognosis and treatment of

cell proliferation conditions especially cancer and ageing  
 Disclosure; Fig 20; 387pp; English.  
 The present sequence represents a human telomerase reverse transcriptase (hTERT) variant from the present invention. The present invention also describes the following methods: (A) determining whether a test compound is a modulator of hTERT, by detecting the change in hTERT recombinant protein or polynucleotide, on administration of the compound; (B) preparation of recombinant telomerase by contacting a protein preparation of hTERT with a telomerase RNA component; (C) detection of the hTERT RNA or protein in a sample by binding a relevant probe to the sample and detecting the complex formed or in the case of RNA detection, amplifying the product with presence of hTERT in the sample; and (D) amplification product with presence of hTERT in the sample; and (E) increasing the proliferation of a vertebrate cell by increasing hTERT expression; and (F) the use of an agent that causes an increase in cell vertebrate cell proliferation to create a medicament that inhibits ageing. A protein preparation of hTERT and the polynucleotide encoding hTERT can be used in the manufacture of medicaments for inhibiting the effect of ageing or cancer. Inhibitors of telomerase activity can be used to treat conditions that are associated with high telomerase activity. A protein preparation of hTERT can also be used in the new methods.  
 Sequence 807 AA;

Query Match 37.2%; Score 2195; DB 1; Length 807;  
 Best Local Similarity 58.3%; Pred. No. 1.1e-194;  
 Matches 462; Conservative 88; Mismatches 191; Indels 52; Gaps 12;

QY 1 MTRAPRCVAVRSLRSRYREVWPLATFVRLGPGRRLLVQPGDKPIYRLVLAQCLVCMHW 60  
 DB 1 MPRAPRCVAVRSLRSRYREVWPLATFVRLGPGRRLLVQPGDKPIYRLVLAQCLVCMHW 60

QY 61 GSQPPADLSEHQVSKELVAVRVQRLCERNERNVLAFCGELLNARGSGPPMAFTSSVR 120  
 DB 61 DARPPAAPSPFQVSKELVAVRVQRLCERNERNVLAFCGELLNARGSGPPMAFTSSVR 120

QY 121 SYLNTVETLRVSGAWMLLSRVGDDLLVYLLAHACALYLLVPPSCAYQVCGSPLYOICA 180  
 DB 121 SYLNTVETLRVSGAWMLLSRVGDDLLVYLLAHACALYLLVPPSCAYQVCGSPLYOICA 180

QY 181 TTDIWPVSASVTRPVGRNFTWRLFLOQIKSSROEAPKPLALPSRGTKRHLSTSTS 240  
 DB 181 ATQARPPPHAS-GPRRLG-----CERAWNHSVREAGVPLGAPGARRRGGASRS 231

QY 241 VPSAKKARCYVPRVEEGP-----HROVLPSPGSKSW-VPSPARSPVEPTAEKDLSSK 292  
 DB 232 LPLKRRPRGAAPERTPVQCGSWAHGPGTRGSDRGFCVVSFAR-----PAEEATSL 286

QY 293 GKVDLSLS-GSVCKHKPSSTLSLSPPRQNAFOLRP-FIETRFHLYSRGQGERLNPSF 350  
 DB 287 GALSCTRSHSPVGRHQHAGPSTSRPPRPWDTPCPVYAEKHELYSSGD-KEQLRPSF 345

QY 351 LLSNLPNLTGARRLVEIFIGSRPRTSPLCRTHLSRRYWMRPLFOOLLVNHAEQY 410  
 DB 346 LLSLRPSLTGARRLVEIFIGSRPMPGTPRRLPRLPQRYWQMRPLFLELLGNAQCPY 405

QY 411 VLLSHCRFRFANQQVDAL-----NTSPHMLDLLKSHSPQVY 452  
 DB 406 GYLLKTHCPLRAA---VTPAAGVCAREKPGQSVAAPEEDTDPRRLVQLLRQHSPPQVY 462

QY 453 GFRLACLVKVSASLWGRHNRERFKNLKFISLGKYGKLSLOELMWMKVEDCHWLS 512  
 DB 463 GFVRACLRPLVPLGWSGRHNRERLNTKFIISLGKHAQLSLQELTWKMSVRDCAWLR 522

QY 513 SPCKORVPAEHRRLREILAKFLHMLMSVYVVELLSRFFYVTTFTQKNRFFFYKKSWS 572  
 DB 523 SPVCGCVPAEHRRLREILAKFLHMLMSVYVVELLSRFFYVTTFTQKNRFFFYKKSWS 582

QY 573 KQSIGVGRHLEVRVRELQSEVRRHQQDTWLAMPICRLFPKPNGLRPVNMYSMT 632  
 DB 583 KQSIGVGRHLEVRVRELQSEVRRHQQDTWLAMPICRLFPKPNGLRPVNMYSMT 642

QY 633 RALGRKQQAQHFTRQLKTLFSLMNTERTKPHLMGSSVLGMNDIYRTWRAFLVRLALDO 692  
 DB 643 RTFREKRAERLTSRVKALFSLVLYERARRPGLGASVLGLDDIHRWRTFVLRVRAQDP 702

QY 693 TPRMYFVKADVTGAYDAIPQGLVEVAVNMIRHSESTYCIROYAVVRRDSQGVHKSFR 752  
 DB 703 PPELYFVKVDVTGAYDTIPQDRLEVIASLIK-PQNTYCVRYAVVQKAHGHVRKAFKS 761

QY 753 QVTTLSLQPYMG 765  
 DB 762 HV-----LRVPVG 769

RESULT 13  
 W97384  
 ID W97384 standard; Protein; 591 AA.  
 AC W97384;  
 DT 14-MAY-1999 (first entry)  
 DE A catalytic telomerase protein.  
 KW Catalytic telomerase; diagnosis; disease; telomerase activity.  
 OS Homo sapiens.  
 PN J11046768-A.  
 PD 23-FEB-1999.  
 PF 01-AUG-1997; 207708.  
 PR 01-AUG-1997; JP-207708.  
 PA (MITU) MITSUBISHI CHEM CORP.  
 DR N-ESDB; X15923.  
 DT New catalytic protein of telomerase of a higher animal and a gene coding it - useful for diagnosis of diseases caused by the change in activity of a telomerase.  
 PS Claim 1; Page 11-14; 18pp; Japanese.  
 CC The present sequence represents a catalytic telomerase protein.  
 CC The products are useful in drug compositions for the diagnosis of diseases caused by the change in activity of telomerase.  
 CC Sequence 591 AA;

Query Match 35.1%; Score 2073; DB 1; Length 591;  
 Best Local Similarity 67.7%; Pred. No. 1.3e-183;  
 Matches 402; Conservative 72; Mismatches 114; Indels 6; Gaps 3;

QY 532 ATFLFWLMDTYVYVQLLSFFYITESTFQKNRFFYRKSVWSKLSIGVROHLERVRLREL 591  
 DB 1 AKFLHMLSVYVVELLSRFFYVTTFTQKNRFFYRKSVWSKLSIGVROHLERVRLREL 60

QY 592 SOEVRHHQDTWLAMPICRLFPKPNGLRPVNMYSMTGTRALGRKQQAQHFTRQLKTL 651  
 DB 61 SEAEVRQREARPALTSRLRIPKPDGLRPVNMIDYVVGARTFREKRAERLTSRVKAL 120

QY 652 FSLMNTERTKPHLMGSSVLGMNDIYRTWRAFLVRLALDOPTPRMYFVKADVTGAYDAIP 711  
 DB 121 FSVLYERARRPGLGASVLGLDDIHRWRTFVLRVRAQDPPELYFVKVDVTGAYDIP 180

QY 712 QGKLVEVAVNMIRHSESTYCIROYAVVRRDSQGVHKSFRQVTTLSLQPYMGQFLKHL 771  
 DB 181 QDRLEVIASLIK-PQNTYCVRYAVVQKAHGHVRKAFKSHVSTLTDLPQTRQFVAHL 239

QY 772 QDSASALNSVVEIOSISNNESSSLDFFLHFLRHSVVKIGDRCYTOCQOIGSSLS 831  
 DB 240 QET--SPLRDVAVIEOSSLINEASSGLFDVFLRFMCHAVRIRGKSYVQCQIGSSLS 297

QY 832 TLLCSLCFGDMENKLPFAEVRDGLLRLVDDFLLVTPHLDQAKTFLSLVHGVPYECMI 891  
 DB 298 TLLCSLCYGMENKLPFAEVRDGLLRLVDDFLLVTPHLDQAKTFLSLVHGVPYECMI 357

QY 892 NLQKTVVNFVEPCTGGAAPYOLPAHCLFPWCGLLDFTOTLEFVCDYSYQAGTSIKSL 951  
 DB 358 NLKRTVYNFVDEALGJGTAFAVQMPAHGLFPWCGLLDFTOTLEFVCDYSYQAGTSIKSL 417

QY 952 TFQSVFVAGKAGTMNRKLLSVLRKCHGLFLDLQVNSLQTCVINYKIFLQAYRFHACVQ 1011  
 DB 418 TFNGRAGRMRKRLFGVLRKCHGLFLDLQVNSLQTCVINYKIFLQAYRFHACVQ 477

QY 1012 LPFDORVRKNTLFFFLGISSQSCCYAILKYNKPGMTLKASG---FPPEAAHWCYQAF 1068  
 DB 478 LPFROQVKNPTFFFLRVLSYDASLCTSLKAKNAGMSLGAKGAGPLPSEAVQMLCHQAF 537  
 QY 1069 LKLAHAHSYIKYCLGLPLRTAQKLLCKRLKPEATWTLKAAADPALSDFTDFTILD 1122  
 DB 538 LKLLTRHRVTVPLVLSLRTAQTLQSLRKLPQTTLTALEAANPALSPDFKTLTD 591

RESULT 14  
 W56109  
 ID W56109 standard; Protein; 564 AA.  
 AC W56109;  
 DT 13-AUG-1998 (first entry)  
 DE Human telomerase reverse transcriptase 63 kDa clone 712562 protein.  
 KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;  
 KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein.  
 OS Synthetic.  
 OS Homo sapiens.  
 FH Key  
 FT Misc\_difference 102  
 FT Location/Qualifiers  
 FT /label= encoded by ARG  
 FT  
 PN GB2317891-A.  
 PD 08-APR-1998.  
 PF 01-OCT-1997; 020890.  
 PR 14-AUG-1997; US-915503.  
 PR 01-OCT-1996; US-724643.  
 PR 18-APR-1997; US-844419.  
 PR 25-APR-1997; US-846017.  
 PR 06-MAY-1997; US-851843.  
 PR 09-MAY-1997; US-854050.  
 PR 14-AUG-1997; US-911312.  
 PR 14-AUG-1997; US-912951.  
 PA (GERO-) GERON CORP.  
 PA (UYTE-) UNIV TECHNOLOGY CORP.  
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,  
 PI Morin GB, Nakamura T, Harley CB;  
 DR WPI; 98-171633/16.  
 DR N-PSDB; V22426.  
 PT Pure and recombinant human Telomerase Reverse Transcriptase and its  
 PT variants - are useful in the diagnosis, prognosis and treatment of  
 PT cell proliferation conditions especially cancer and ageing  
 PS Example 1: Fig 68; 387pp; English.  
 CC The present sequence is a human telomerase reverse transcriptase (hTERT)  
 CC clone protein from the present invention. The present invention also  
 CC describes the following methods: (A) determining whether a test compound  
 CC is a modulator of hTERT, by detecting the change in hTERT recombinant  
 CC protein or polynucleotide, on administration of the compound; (B)  
 CC preparation of recombinant telomerase by contacting a protein  
 CC the hTERT RNA or protein with a telomerase RNA component; (C) detection of  
 CC sample and detecting the complex formed or in the case of RNA detection,  
 CC amplifying the product and correlating the presence of complex or  
 CC amplification product with presence of hTERT in the sample; and (D)  
 CC increasing the proliferation of a vertebrate cell by increasing hTERT  
 CC expression; and (E) the use of an agent that causes an increase in cell  
 CC vertebrate cell proliferation to create a medicament that inhibits  
 CC ageing. A protein preparation of hTERT and the polynucleotide encoding  
 CC hTERT can be used in the manufacture of medicaments for inhibiting the  
 CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
 CC used to treat conditions that are associated with high telomerase  
 CC activity. A protein preparation of hTERT can also be used in the new  
 CC methods.  
 SO Sequence 564 AA;

Query Match 25.5%; Score 1506.5; DB 1; Length 564;  
 Best Local Similarity 58.9%; Pred. No. 4.5e-131;  
 Matches 297; Conservative 52; Mismatches 82; Indels 73; Gaps 3;  
 QY 539 MDYVVLVLLSFFYITETQKNLFFYRKSVWSKLQSIGVROHLERVLRELSEAEVRH 598  
 DB 1 MDYVVLVLLSFFYITETQKNLFFYRKSVWSKLQSIGVROHLERVLRELSEAEVRH 598

DB 1 MSYVVVLLRSFFYITETQKNLFFYRKSVWSKLQSIGVROHLERVLRELSEAEVRQ 60  
 QY 599 HODTWLAMPICRLRFIPKPNGLRPVNNMSYSMTGTRALGRKQAOHFTORLTLFSLNLYE 658  
 DB 61 HREARPALTLRLRFIPKPNGLRPVNNMDYVVGARTFRREKXAERLTSRVKALFSLNLYE 120  
 QY 659 RTKHPLMGSSVLGNDIYRTWRAFLVRALDQTPRMFYFKADVGTGAYDAIPQCKLVEV 718  
 DB 121 RARRPGLLGASVLGLDDIHRAWRTFVLVRADQPPPELYFVKVDVGTGAYDTIPQDLTEV 180  
 QY 719 VANMIRHSESTYCIROYAVVRDSSOGQVHKSFRRQVITLSDLPYMGQFLKHLQSDASA 778  
 DB 181 IASIIK-PQNTYCVRYAVVQAAAMGTSAEPSRA----- 214  
 QY 779 LNSVVISQISISMNESSSLDFELHFLRHSWVKIGRCYTCOCGIPQGSLSLTLCSLC 838  
 DB 214 -----TSYVQCGIFQGSLSLTLCSLC 236  
 QY 839 FGDMEKLPFAEQVQDGLLLRFVDDFLVTPHLDQAKTFLTLVHGVPEYGCMTNLQKTVV 898  
 DB 237 YGDMENKLPFAIRKRDGLLLRVDDFLVTPHLDQAKTFLTLVHGVPEYGCMTNLQKTVV 296  
 QY 899 NFPVEPTGLGGAAPYQLPAHCLFPWCGLLDLDTQTFLEVCDSYGAQTSIKTSLTFQSVFK 958  
 DB 297 NFPVEDEALGGTAFVQMPAHGLFPWCGLLDLDTQTFLEVCDSYGAQTSIKTSLTFQSVFK 356  
 QY 959 AGKTRNKLKLVRLKCHGLFLDLQVNSLTQVCINIKYKIFLQAYRFHACVIOLPDQV 1018  
 DB 357 AGRNRRKLFGLVRLKCHGLFLDLQVNSLTQVCINIKYKIFLQAYRFHACVIOLPDQV 416  
 QY 1019 RKNLTFELGISSQSCCYAILKV 1042  
 DB 417 WKNPHE-----SCASLTRL 431

RESULT 15  
 W47001  
 ID W47001 standard; Protein; 538 AA.  
 AC W47001;  
 DT 13-AUG-1998 (first entry)  
 DE Glutathione-S-transferase and hTERT fusion protein 1.  
 KW Human; telomerase reverse transcriptase; hTERT; TRT; diagnosis;  
 KW prognosis; cell proliferation; cancer; ageing; ribonucleoprotein.  
 OS Synthetic.  
 OS Homo sapiens.  
 FH Key  
 FT Region  
 FT Location/Qualifiers  
 FT 1..220  
 FT /note= "glutathione-S-transferase fragment"  
 FT Misc\_difference 221..226  
 FT /note= "thrombin cleavage sequence"  
 FT Misc\_difference 227..231  
 FT /note= "heart muscle protein kinase recognition sequence"  
 FT Misc\_difference 232..236  
 FT /note= "residues introduced by cloning"  
 FT Region  
 FT 237..538  
 FT /note= "hTERT protein fragment"  
 FT GB2317891-A.  
 PN 08-APR-1998.  
 PD 01-OCT-1997; 020890.  
 PF 14-AUG-1997; US-915503.  
 PR 01-OCT-1996; US-724643.  
 PR 18-APR-1997; US-844419.  
 PR 25-APR-1997; US-846017.  
 PR 06-MAY-1997; US-851843.  
 PR 09-MAY-1997; US-854050.  
 PR 14-AUG-1997; US-911312.  
 PR 14-AUG-1997; US-912951.  
 PA (GERO-) GERON CORP.  
 PA (UYTE-) UNIV TECHNOLOGY CORP.  
 PI Andrews WH, Cech TR, Chapman KB, Harley C, Lingner J,  
 PI Morin GB, Nakamura T, Harley CB;  
 DR WPI; 98-171633/16.  
 DR Pure and recombinant human Telomerase Reverse Transcriptase and its

Mon Aug 16 10:21:26 1999

PT variants - are useful in the diagnosis, prognosis and treatment of  
PT cell proliferation conditions especially cancer and ageing  
PS Example 6; Page 224; 387pp; English.

CC The present sequence represents a fusion protein from an example  
CC of the present invention which describes human telomerase reverse  
CC transcriptase (hTERT). The present invention also describes the  
CC following methods: (A) determining whether a test compound is  
CC a modulator of hTERT, by detecting the change in hTERT recombinant  
CC protein or polynucleotide, on administration of the compound;  
CC (B) preparation of recombinant telomerase by contacting a protein  
CC preparation of hTERT with a telomerase RNA component; (C) detection  
CC of the hTERT RNA or protein in a sample by binding a relevant  
CC probe to the sample and detecting the complex formed or in the case of  
CC RNA detection, amplifying the product and correlating the presence of  
CC complex or amplification product with presence of hTERT in the sample;  
CC and (D) increasing the proliferation of a vertebrate cell by increasing  
CC hTERT expression; and (E) the use of an agent that causes an increase in  
CC cell vertebrate cell proliferation to create a medicament that inhibits  
CC ageing. A protein preparation of hTERT and the polynucleotide encoding  
CC hTERT can be used in the manufacture of medicaments for inhibiting the  
CC effect of ageing or cancer. Inhibitors of telomerase activity can be  
CC used to treat conditions that are associated with high telomerase  
CC activity. A protein preparation of hTERT can also be used in the new  
CC methods.  
SQ Sequence 538 AA;

Query Match 18.5%; Score 1092.5; DB 1; Length 538;  
Best Local Similarity 71.2%; Pred. No. 1.1e-92;  
Matches 215; Conservative 26; Mismatches 58; Indels 3; Gaps 1;  
QY 824 IPQSSILTLCLSCFGDMENKLFPAEVRDGLLRVFDDELLVTPHLDQAKFLSTLVHG 883  
DB 237 IPQSSILTLCLSCYCGDMENKLFAGIRRDGLLRVLDVDELLVTPHLDQAKFLSTLVHG 296  
QY 884 VPEYGCMINLQKTVVNFPEFTGLGGAAPYOLPAHCLFPWCGLLDQTGLEVCDSGYA 943  
DB 297 VPEYGCVINLRKTVVNFPEFTGLGGAAPYOLPAHCLFPWCGLLDQTGLEVCDSGYA 356  
QY 944 QTSIKTSLTFSOFKACKTMENKLLSVLRKCHGLFLLDQVNSLQTVCIINIKYIFLIQAY 1003  
DB 357 RTSIRASVTENRFRAGNRMRKLFGLVLRKCHSLFLLDQVNSLQTVCIINIKYIFLIQAY 416  
QY 1004 RFHACVQLPFDQVRKKNLTFFLGIISQASCCYAILKVNPGMTLKASG---FPPEAA 1060  
DB 417 RFHACVQLPFEQVQWKNPTFFLRVISTASLCYSILKAKNAGMSLGAKGAGLPSEAV 476  
QY 1061 HMLCYOAFLLKLAHSVIYKGLGLPRTAOKLGLCKLPKPEATMTILKAAADPALSTDFQTI 1120  
DB 477 QWLCHOAFLLKLRHRTVTVPLGLSLRTAQQLSKRLPGTTLTALAAANPALPSDFKTI 536  
QY 1121 LD 1122  
DB 537 LD 538

Search completed: August 13, 1999, 21:34:53  
Job time: 999 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 1999, 21:18:21 ; Search time 16.13 Seconds  
(without alignments)  
2786.948 Million cell updates/sec

Title: US-09-042-460-2  
Perfect score: 5901  
Sequence: 1 MTRAPRCPAVRLRSRYRE.....TILKAADPALSTDFQTILD 1122

Scoring table: BLOSUM62

Searched: 122810 seqs, 40065486 residues

Database : PIR\_60: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	341	5.8	884	2 S53396	telomerase catalyt
2	131	2.2	778	2 I38487	tastin - human
3	125	2.1	1529	2 T02730	probable reverse t
4	121	2.1	2088	2 E71436	hypothetical prote
5	115.5	2.0	2157	2 S71461	proline-rich prote
6	111.5	1.9	1415	1 E0BEGA	immediate-early pr
7	111.5	1.9	2606	2 T03159	large tegument pro
8	110	1.9	907	2 A24938	hypothetical t2 pr
9	109.5	1.9	1214	2 T00356	hypothetical prote
10	108.5	1.8	351	2 S50754	hypothetical prote
11	108	1.8	880	2 S49627	regulatory protein
12	107	1.8	1456	2 T01397	LTR transposon pol
13	106.5	1.8	1256	2 C71436	probable resistanc
14	106	1.8	1115	2 A45761	Ca2+-transporting
15	106	1.8	584	2 S40013	hypothetical prote
16	105.5	1.8	3898	2 S57437	genome polyprotein
17	104.5	1.8	916	2 A38418	Jockey protein 2 -
18	103.5	1.8	536	2 E70066	hypothetical prote
19	103	1.7	508	2 S74977	hypothetical prote
20	102.5	1.7	763	1 A40831	gag-akt polyprotei
21	102.5	1.7	1094	2 T00814	probable zinc-fing
22	102.5	1.7	1556	2 S59393	probable membrane
23	102.5	1.7	781	2 A37956	sulfate permease I
24	102	1.7	908	2 S07649	gene coI intron 1
25	102	1.7	850	2 S56015	gastric mucin - hu
26	101.5	1.7	884	2 T01041	hypothetical prote
27	101	1.7	994	2 I49276	c-mer tyrosine kin
28	101	1.7	1004	2 T00795	tRNA-processing pr
29	100.5	1.7	1684	2 T02632	hypothetical prote
30	99.5	1.7	360	2 S18458	tubulin beta chain
31	99.5	1.7	810	2 S49744	AMP deaminase (EC
32	99.5	1.7	968	2 T00322	hypothetical prote
33	99.5	1.7	528	2 I47141	gastric mucin (clo
34	99	1.7	1680	2 T01367	hypothetical prote
35	99	1.7	1228	2 S59681	probable membrane
36	99	1.7	840	2 S69204	pheromone response
37	99	1.7	1621	2 S62356	TRP-185 protein -
38	99	1.7	646	2 JC5583	85K SH3 domain-con
39	98.5	1.7	3898	2 S58295	polyprotein, - hog

40 98.5 1.7 2163 2 S50675  
41 98.5 1.7 5262 2 T03454  
42 98 1.7 1239 2 I49705  
43 97.5 1.7 916 2 J70396  
44 97 1.6 1174 2 A43736  
45 97 1.6 701 2 S64737

ALIGNMENTS

RESULT 1  
S53396  
telomerase catalytic chain EST2 - yeast (Saccharomyces cerevisiae)  
N:Alternate names: protein L8543.12; protein YLR318W  
C:Species: Saccharomyces cerevisiae  
C:Date: 05-May-1995 #sequence\_revision 01-Sep-1995 #text\_change 13-Sep-1998  
C:Accession: S53396  
R:Du, Z.  
submitted to the EMBL Data Library, February 1995  
A:Description: The sequence of S. cerevisiae cosmid 8543.  
A:Reference number: S53396  
A:Accession: S53396  
A:Molecule type: DNA  
A:Residues: 1-884 <DUZ>  
A:Cross-references: EMBL:U20618; NID:G2258165; PID:G662136; MIPS:YLR318W  
A:Experimental source: strain S288C (AB972)  
C:Genetics:  
A:Gene: SGD:EST2  
A:Cross-references: SGD:S0004310; MIPS:YLR318W  
A:Map position: 12R

Query Match 5.8%; Score 341; DB 2; Length 884;  
Best Local Similarity 19.8%; Pred. No. 6e+18;  
Matches 208; Conservative 148; Mismatches 355; Indels 338; Gaps 40;  
QY 33 PEGRRLLVQPGDPKYYRTLVLAOCVCMHWGSPPPADLSFHQVSSLLKELVARVORLCERN 92  
DB 45 PNRKIALP-----CL-----PGDLSHKAV-----IDHCIIYLLTGEL 77  
QY 93 ERNVLAFGFELL-NEARGGPPMAFTSVRSYLPNTVITLRVSGAWMLLSRVGDDLLVY 151  
DB 78 YNVVTFGFKIARNEDVNSLFCHSANV-----VTLLGAAMKMFSLVGTAFVD 129  
QY 152 LLAHCAALYLLPPSCAYQVCGSPLYQI-----CATDTP-----SVSASYRTPRVPGR 200  
DB 130 LLIN--YTVI-----QFNQOFFQIVGNRCNEPHLPKQVQRSSSSSATAAQIKOLTE 180  
QY 201 NETNLRFLOQIKSSSRQBPAPKPLALPSRGTKRHLSTSTSVPSAKKARCYPVPRVEGPH 260  
DB 181 PVTNQFQFHK-----LNINSSSF-----FPY 201  
QY 261 ROVLPTPSGKSWPSPARSPVPTAEKDLSSGKVDLSGSGVCKKHPSPSTSLSPPR 320  
DB 202 SKILPSS-----SSIKKLTDL-----REAIFPTNLVKIP- 231  
QY 321 QNAQLAPFETRHFYLSRGDQGERLNPSFLSLNQLPNTGTARRLVEIIFLGSRPRTSGP 380  
DB 231 -----QRL-----KVRINLT----- 241  
QY 381 LCRTHRLSRRYQWRPLFOQLLYNHAECQVYRLLRSHCRFRTANOQVTDALNTSP--H 437  
DB 241 -----LQKLLRKHRLNYSTLNSIC-----PPELGT 267  
QY 438 LMDLLRL--HSSPWQYGFELRACLCKVWSASLWGTSHNERREFKLNKFIISGKVGKLSL 495  
DB 268 VLDSLHLSRQSPRERYLKFIIIVLOKLLPQEMFGSKNGKIKNLNLNLLSLPLGYPFP 327  
QY 496 QELMKMKVDECHWLSSSGKGRVPAAEHRLR--ERILATFLWMDYVYVQLRSFFYI 553  
DB 328 DSLKLLKRLKDFWELFISD-----IWFTHNFENLNQLAICFTISLWFLRQILPIKIOTFFYC 383

Mon Aug 16 10:21:27 1999

554 TE-STFQKNLRFYKRSVWSKQSIQVROHLRVLRLSREBVEV--RHHDQTLWAMPIC 609  
 384 TEISSTV---TIVFRDHTWKNLITPFIVEFKYI---LVENVCRNHSYTLNFSNHS 436  
 610 RLRETPKP--NGLR-----PIVNMYSMGTRALGRKKAQHTORLKLTLFSLMNLVTKHP 663  
 437 KMLIPKSNNEFRILAIAPCGADEEETIYKENHNAIQTKI-----LEYLRNKR 490  
 664 HLMGSSVLGMNDIYTWRALVLRV--RALDQTPRMVFKADVTGAYDAIPQGLVEVVAN 721  
 491 TSF-TKIYSPTQIADRIKEFKORLLKFNVLPELIFMKFDVKSCYDIPMECMRILKD 549  
 722 MIRHSESTVICROY-----AVVRDSQGVHKSFRQVTTLSLDLOPYMGFLK 769  
 550 ALKNENGFFVRSQYFNTGVLKFNVNASRPVPELYIDNVRTV----- 598  
 770 HLODSDASALRNSVIEQISINNESSSLFDFLHLRHSVVKIGDCYTOCQIPOGSS 829  
 598 HLSNQDV-----INVMEIF-----KTALWVEDKCIYREDGLFQSS 635  
 830 LSTLLCSLCEGDM---ENKLEAEVQDGLLRFVDDFLIVTPLHDOAKTFSTLVHGVPE 886  
 636 LSAPVDLVDDLEFYSEKASPSQDTLLKLADDFLIISTDQOVINIKKILAMGSGFQK 695  
 887 YGCMINLQKTVVNPVPEPTLGGAAPIQLPAHCLFPWCGLLLDTQTLVFCDSYGAQTS 946  
 696 YNAKANRDKIL-----AVSQSDDDTVIOFCAMHIFVRELEVKHSS----- 738  
 947 IKTSLTQSQVFKAGTKMRNKLVLRLKCHGLFDLQVNSLQVNCINI-----YK 996  
 738 --TMNNEHRSKSGKIFRSIALFNTRISYKIIDNLNSTNLVLMQIDHVVKNISECYK 795  
 997 -----IFLLQAYRFFHACVIQPFDDRV 1018  
 796 SAERDLSINVTQNMQFHS-----FLQRI 818

RESULT 2  
 138487  
 tatin - human  
 C:Species: Homo sapiens (man)  
 C:Date: 16-Feb-1996 #sequence\_revision 16-Feb-1996 #text\_change 16-Feb-1996  
 C:Accession: 138487  
 R:Fukuda, M.N.; Sato, T.; Nakayama, J.; Klier, G.; Mikami, M.; Aoki, D.; Nozawa, S.  
 Genes Dev. 9, 1199-1210, 1995  
 A:Title: Trophinin and tatin, a novel cell adhesion molecule complex with potential inv  
 A:Reference number: 138487; MUID:95278733  
 A:Accession: 138487  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-778 <RES>  
 A:Cross-references: EMBL:U04810; NID:905355; PID:905356

Query Match 2.2%; Score 131; DB 2; Length 778;  
 Best Local Similarity 22.4%; Pred. No. 0.061;  
 Matches 107; Conservative 44; Mismatches 167; Indels 160; Gaps 26;  
 2 TRAPCPAYRLSRLSYREVWPLATEVRLGPGRRLVQGPDKYIRLVA-----QC 54  
 358 TPNPSTPRVQQ-----AQLRGVSPQS-----CSEDPAFPWEQVAVRLFDQESC 401  
 55 LVCMHNGSQPP---PADLSFHQVSSKLKELVAR---VVORLCERNERNVLAFGLLNEAR 108  
 402 IRSLESGKPPVATPSGPHSNRTPSLOEVKIORIGTLQQLLRQEVGLVGGCVPLN--- 459  
 109 GCP--PNAFTSSVRSYLPNTVIETLRVSGAMLLLSRVGDDLLVYLLAHCALYLLVPPSC 166  
 459 GGSLLDWELOPLTLTSRLNATEHNSGT-----SHLPGLLKHSGL-----PKPC 504  
 167 AYQVCGSGPLVOICATTDIWP-----SVSASVYRTPRPVGRNFTN-L 205  
 505 LPEECGEP--QPCPPAEPGPPAFCRSEPEIPEFSLOEQLEVPEPYPAP--RPLESCC 560

206 RFLQIKSSRQE-----APKPL-----ALPSRGTKRHLSTSTSVPSAKK 246  
 561 RSEPEIPSSQEQLEVPPECPAEPRLSESYCRIEPEISSRQEOL-----EVPE--- 613  
 247 ARCYVPVVEGPHQVLPPTSGKSWVPSPARSEVPTAEKDLSSKGKVDLSLSGVS-- 305  
 613 ----PCPPAEPGP---LQPSQGGSGPPGC-----PRVELGASEPCTLEHRSLESLLPP 660  
 305 CCKH-KPSTSL-----SPRONAFOLRDIETRHFLYSRGQGERLNPFSLLSN 354  
 661 CCSQWAPATTSILFSSQHPCLASPPICSLQSLRP-----PAGQAG-----LSN 703  
 355 LOPNLTGARLV-----EIFLGSRPRTSGP--LCFTHRLSRRYQ 393  
 704 LAPRTLALRESLKSCLTATHCHEARLDDECAFYTSRASPSGPTRVCTNPVATLEWQ 761

RESULT 3  
 T02730  
 probable reverse transcriptase T914.6 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 24-Mar-1999  
 C:Accession: T02730  
 R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.;  
 submitted to the EMBL Data Library, August 1998  
 A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.  
 A:Reference number: Z14710  
 A:Accession: T02730  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1529 <ROU>  
 A:Cross-references: EMBL:AC005315; NID:g3461834; PID:g3461840  
 C:Genetics:  
 A:Map position: II  
 A:Introns: 428/1; 532/3  
 A:Note: T914.6

Query Match 2.1%; Score 125; DB 2; Length 1529;  
 Best Local Similarity 19.2%; Pred. No. 0.48;  
 Matches 209; Conservative 147; Mismatches 341; Indels 394; Gaps 58;  
 142 SRVGDDLLVYLLAHCALYLLVPPSCAYQVCGSPYQICATIDIPWSASVYRTPRVGRN 201  
 220 SETGYDAMVEY-----KYPWLPSPRSCSKWGHIOEVCLTR---PSPQLSTPT----- 266  
 202 FTLNRLQIKSSSRQEPAP-----KPLALPSRGTKRHLSTSTSVPSAKKARCYVP--PRV 255  
 266 -----EIEETEDKTEPPLMKRPLEILSKSPSATLTKTLNGDSHTQKV---PMKNPTV 314  
 256 -----BEGPHRQVLPPTSGKSWVPSPAR-----SP-----EYPTA 285  
 315 LQNKGEVAEENEENLKDEGPMHTVSPAKVGRQSPQOEYVNNHASPSPRFDVLAVEDTA 374  
 286 EKDLSSKGVSDLSLGSVCCKHKPSTSLSPPRQNAFOLRPEI-----ETRHFLYSRGD 341  
 375 DTNGHEGEIVVLQ-----QLDNDSV-----GVQGTPTPIPRISKQHKVVKSSA 422  
 342 GQE-----RLNPSFLLSNL-----QPNLTGARLVVEI 368  
 423 NQNTKGIWVWRDNARLTVPFKSSQMITCSILLGKEEEFFCFSFIYASNFVEER---I 479  
 369 IFLGSRPRTSGPLCRTHRLSRRYQWRPLFOQLLVNHAECQVRLLRSHCRFRFTANQOVT 428  
 480 LWEDIRSHHDSPLIR-----NKPWILCGDFNEILGEGHSNY----- 517  
 429 DALNTSPHPLMDLLRLHSSPWQVYGFRLACLCKVVSASLWGTNRHNRERFKNLKFFISLG 488  
 517 DNSPYTPPGMRD-----FQBIGRL-----MLEAATGGR---KPF-----KEVNV- 554  
 489 KYGLSLQELMKWKVEDCHWLRSSP-----GKDRVPAAE 523



Db 554 -----LTKLPQFLPVVSHWASAPLYVSTALYRESKLLKTLKPHILRELKGLGDLP 607  
QY 524 HRLRE-RIL-----ATFLWMDTVVQLLRSFFYIT-----ESTFOK----- 561  
Db 608 KRTREAHILCEKQATTLANPSOETIAEELKAVTDMTHLSELEGFLKQSKLHWNVGD 667  
QY 561 -NRLFFYRKSWKSLQSIGVROHLERVLREL-----SDEEVRHHODTWLAMPICR 610  
Db 668 GNSYFHKAQVRKM-----RNSIREIRGPNATLQTSSEEKGEAREFN----- 713  
QY 611 LRIPKPN-----LRPIVMSYMGTRALGRKQKQAHFTQRLTKTLFSLMNYERTK 661  
Db 713 -EFLNQSGDFHGISVEDLNLMSYRCSVTDQNLITREVTGEEIQ--KVLFAFN-----NK 766  
QY 662 HPHLMGSSVLGMNDIYR-TWRA-----FVLVRALDQTPMYFVKADVTGAYDAIPOGLV 716  
Db 767 SPQPDGYT-----SEFFKATWSLTPDFIAAIQS-----FFVKG-----FLPKGLNA 808  
QY 717 EVVANNIRHSES-----TYCIRQYAVVRDSDQGVHKSFRROVTTLSDLQPYMGOF 767  
Db 809 TILALIPKDEATEMDYRIPISCNVLYKISKILAN-----RLKLLPSFILQNSAF 862  
QY 768 LKHQSDASALRNSV-----IQSISMNESSSLFD-----FFLHFL----- 807  
Db 863 VKERLLMENVLLATELVKDYHKSVPTRCANKIDISKAFDSVQWQFLLTLEALNPPET 922  
QY 807 RH-----SVYKIGDRC--YTQCOGIPQSSLSLLSCFCGDMENKL-FAEVOR 852  
Db 923 RHWIKLCISTAFSVQVNGELAGFFGSSRGLRGCCALSPYLFVICHNVLSHMIDEAAVHR 982  
QY 853 D-----GLL-LRFVDDFLL-VTPHLDQAKTFLSLTVHGVPEYGCWMLNQ----- 896  
Db 983 NIGYHPKCEKIGLTHCFADDLVDFVGHQWSTEGVINFKFAGRSGLQISLEKSTIYL 1042  
QY 896 -----TVNFPVEPTGLGGAAPYOLPAHCLPFWCGLLDQTLQVDFCDYGYAQT 945  
Db 1043 AGVASDRVQTSUSFFPANG-----QLPVRYL-----GLPLTKQMT-ADYSPLE- 1089  
QY 946 SIKT-----SLTFOSVFKAGTMRKLLSVLRKCHGLFLDLQVNSLQTCVINYKFLQOA 1002  
Db 1089 AVTKISSWTARSLSVAGR-----LALLNSVIVSIAN-FWMSA 1125  
QY 1003 YRFHA-CVIQL 1012  
Db 1126 YRLPAGCIREI 1136

RESULT 4  
E71436  
Hypothetical protein - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
A:Variety: Columbia  
C:Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998  
C:Accession: E71436  
P.: Bent, E.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirksen, P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenecker, T.; Pohl, T.M.; Terry, N.; Gielan, T.; Hempel, S.; Kottler, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech, A.; Moeres, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans, C.; Chludzisz, N.  
A:Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana  
A:Reference number: A71400; MUID:98121113  
A:Accession: E71436  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-2088 <BE>  
A:Cross-references: G5:297342; NID:g2245031; PID:e327020; PID:g2245044  
C:Genetics:  
A:Map position: 4COP9-4G3845

Query Match 2.18; Score 121; DB 2; Length 2088;

Best Local Similarity 19.9%; Pred. No. 1.5;  
Matches 151; Conservative 82; Mismatches 251; Indels 276; Gaps 37;  
QY 88 LCERNERNVLAFCFFELLNEA-----RGGPPMAFTSSVRSYLPNTV----- 128  
Db 1293 LSEKRRHIVETGLTLLTQASVPREYWPYFAAAV--YLINRMPVLSMESPFQKLFGS 1350  
QY 128 ---IETLRVSGA---WM-----LLSRVGGDLLVLLAHALY----- 160  
Db 1351 KPNYERLAVFGCLCPWLPYPYTHNKLERSRCVFLGSLTQTAYLCEDEVEHKRLYTSRH 1410  
QY 160 -----LLVP-----PSCAYQVCGSPY-QICATTDIWPVSASYSYRTPRPVGRNFT 203  
Db 1411 VVFEASFPFSLNTSQSLPTVTFEQSSPLVPTPLSSSSVLPSCLS--PCTVLHQOQP 1468  
QY 204 NURELQIKSSSRQAPKPLAPSRGTXRHLSTSTSVPS-----A 244  
Db 1469 PVTNSPHSQPTTSPAPLS-PHRSITMDFOVPOPTAPENGEPEPEAQSPPIGSLNPT 1527  
QY 245 KKARCYPVPRVEEGPHROVLPPTSGKSWPSPARSPEVPTAEKDLSSKGKYSGLSLSGSV 304  
Db 1528 HEAFIGPLPNRNPNTNIEPT-----PAPHKPKVPTTTTTPNRTTVDAS----- 1576  
QY 305 CCKHXPSTSLSPRONAFOLRPIETRFLYSGDQ--ERLNPFSLLSNLQNLGTAR 363  
Db 1576 ---HOPTA-----POON-----QHNKTRAKNNIKKPNTKFSLTATLPNRS-- 1614  
QY 364 RLVEIFIGSRPSGGLCRTHLSRRYWO-----MRPLFQQLLVNHA 406  
Db 1614 -----PSEPTNVQALKDKKRFAMSDFEAQOQNHHTWDLVPHESQLLVG-- 1659  
QY 407 ECVYVRL-----RSHCRFTANQO--VTDALNTSPPHLMDLLRL-----HSSP 448  
Db 1659 -CKWVEKLYLPNGAIDKYKARLVAKGQVGYDAETFSVPIKSTIRLVLDVAVKKD 1717  
QY 449 WQVYGLRACLCVVSASLWGTNRINERFFKNKKFISLGKYGKLSLOELMKMKVEDCH 508  
Db 1718 WEIKOL-----DYNNAFLOGLTTEE-----V 1738  
QY 509 WLRSPG---KDRVPAAEHRLRERILATFLWMDTVVQLLRSFFYI-TESTFQKNRLF 564  
Db 1739 YMAQPPGFIDKDR-PTHVCLRLKAIYG--LQAPRAYMELKQHLFNIGFVNSLSDASLF 1795  
QY 565 FYKSVSKLSIGVROHLERVLRLRELSQEVRRHODTWLAMPICRLRFPKPNGLRPV 624  
Db 1796 IY-----WSDKSSID-----AVLTSIAE-----RF-----SIKDPT 1821  
QY 625 NMSYMGTRALGRKQAHFTOR--LKTFLSMLNYERTK-----HPHLMGSSVLGMN 674  
Db 1822 DLHYFLGIEAT-RTKQGLHLMQRYIKDLLAKHNMADAKPVLTPLTSPKTLTHGGTKLN 1880  
QY 675 DI--YRTWRAFLVRALDQTPMYFVKADVTGAYDAIPQ 712  
Db 1881 DASEYRS-----VGSIQIYLAFTRPDIAYAVNRLSQ 1911  
RESULT 5  
S71461  
proline-rich protein Rad9 - inky cap (Coprinus cinereus)  
C:Species: Coprinus cinereus  
C:Date: 09-Dec-1997 #sequence\_revision 09-Dec-1997 #text\_change 26-Feb-1998  
C:Accession: S71461; S78194  
R:Seitz, L.C.; Tang, K.; Cummings, W.J.; Zolan, M.E.  
Genetics 142, 1105-1117, 1996  
A:Title: The rad9 gene of Coprinus cinereus encodes a proline-rich protein required for  
A:Reference number: S71461; MUID:96271528  
A:Accession: S71461  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-2157 <SEI>  
A:Cross-references: EMBL:U34998; NID:g1353389; PID:g1353390  
A:Accession: S78194  
A:Molecule type: mRNA

A: Residues: 1-2157 <SET>  
A: Gene: rad9  
C: Genetics: rad9  
A: Introns: 65/1; 151/2; 433/3; 472/1; 685/3; 844/3; 878/2; 1080/1; 1114/3; 1182/2; 1246/2; 2054/3  
C: Keywords: DNA repair; melosis  
F: 607-616/Region: nuclear location signal

Query Match 2.0%; Score 115.5; DB 2; Length 2157;  
Best Local Similarity 18.3%; Pred. No. 4.3; Mismatches 439; Gaps 54;  
Matches 213; Conservative 155;

QY 70 SFHQVSS-LKELV-----ARVQRLCERNERNVLAF-GPELLNEARGGPPMAFTSSVRS 121  
Db 884 SIRTVSALLMQLVQTSNDVIRINARLEKERQNALAKROESISDLNGQP-----KSDPE 938  
QY 122 YLPNTVIETLRVSGAM-----LLLSRVG-----DDL 148  
Db 939 FLDNIDMEERLYGGGLESATKAATTIIFLNSRAGRKTTKNSNEAYRAIFDNLDL 998  
QY 149 LVYLHACALYLLVPPSCAVCGSPLYQICATTDIWPVSASYPTRPVGRNFTNL--- 206  
Db 999 LVVL-----YWPDPAAASLLLNIAISM 1019  
QY 206 ---RFLQOIKSSRP--QAPKPLALPSRGTKRHLSLTSTSVPSAKKARCYVPVPRVEEGPH 260  
Db 1020 FVWSSLDVKSNAQIDTNAKSMALD-----HLGVIAAIRSS-----ILKVQKDEGTS 1069  
QY 261 RQVLPTSGSKSWPSP-----ARSPEV-----PTAEKDLSSKKVSDLSLGSVC 305  
Db 1070 YRGLKPLDEVDRQSPVEAIRQVHGRSPDVAHAHLCKRSTEDQAVESARELTAAILGHELG 1129  
QY 306 CKHKPSSTLSLPPRQNAFOLRIETRHLYSRGDOERLNPFLSNLQPNLTGARR- 365  
Db 1130 ASLRVNWLDHPQDEDLDR-----DSSKAL--SFGL-KLKTALRGWKD 1173  
QY 365 -LVEIIFLGSRPRTSGPLCRTHRLSRRYWQMRPLFOQLLVNHAECQVRLRSHCRPTA 423  
Db 1174 PATDVFVGISQEEVS-----RIDRLSE-----EIGTIQSLR----- 1205  
QY 424 NQVTDALNTSPHMLDLRLHSSPQWYVGLRA-----CLCKVYS--ASLWGTGRHNERRF 477  
Db 1205 -----NSFQPLNVLISALDAP---VIFWRTKALRQCIQVTSATILGTA-SVRQG 1252  
QY 478 FNLKKTFTSLGKYGLSLQELMWMKVEDCHWLRSPP-----GKORVPAE----- 524  
Db 1253 IEN-----HLLDSSPAVRDAAVELICKYIMDSPEVAGN 1285  
QY 524 --HRLRERILATFLWMDTYVQLLRSEFFYITESTFKNLFYRKSVMWSKLOSIGVRQ 581  
Db 1286 YVQKTAERMADTGL--AVRKRVIKLLKSYGVVIDDAQRKIDI----- 1326  
QY 582 HLVRRLRELQOE-----VRHODTWLAMPICRLRPIKPNGLRP----- 623  
Db 1326 --SARILVRNVEDDGVKDLAKMTLEELWFPp-----LPPSGMKVKPTSSSNPNQDKA 1377  
QY 623 ---IVNMSYSMTGRALGRKQA-----QHFTORLKT----- 652  
Db 1378 ALLSKVAIIMGTAANFRDQRSPLEMDLHKIISDKEGDEAASLHORVABICETLIDGLVDA 1437  
QY 652 -----FSMLNERTKH-----PHLWSS-----VLGMNDIVRTWRAF 683  
Db 1438 TDLPGFTIINCIRTIHLFTAAVPSILPGTHASTLLPYLKNASTTEELTSLDFLLKIFRA- 1497  
QY 684 VLVRALDQTPRMFEVK-ADVTGAYDAIPQGLVEVAMNIRHSESTYCIROYAVVRDS 742  
Db 1497 --SIPHMPKTPQIRAGIADVAALILKPFQGV-----NILQEAAGCMC-----AVVR--- 1543  
QY 743 QGVHKSFRQVTTLSQLOPYMGQFLKHLQSDASALRNSVIEOSISWNNESSSLDFDF 802  
Db 1543 --HLTHDFKRLINLLKGCNARLLSYLRH-----PPTKQLNNV-----ESKTLMLLF 1587

QY 803 LHELRLHSHVVKIGDRG-----YTQCQGIPOGSSLSLTLCSL----- 838  
Db 1588 I-----VALLGEHCNFORLREQDPLADIDSTIQSGVMEHIYFTLLRIYDKDFADIR 1641  
QY 838 -----CFG-----DMENKLFAYQORDG--LLLRFDVDFLLVTPHLDQAK 874  
Db 1642 PRILOCLGFLFRAQPTLMTKEESAINDAIFASEEEGRARLLKIMQDFLISESEKHSK 1701  
QY 875 TFLSTLVHGVPEYCGMINLQKTVVNFVPEPTLGGAAPIQLPAHCLFPWCGLLLD---TQ 931  
Db 1702 E-----KESAKNKANKATDVNMEELVGTDFADSGVSSAIVQRYLSHILDAALSQ 1752  
QY 932 TLEVECDYSGAQTSTIKTSL--TFQS--VFRAKGTMRNKLLSVLRKCHGLFGLDQVNSL 987  
Db 1753 NSQIQMAAIDVLTFTIKOGLAHTLOSFPVITALETSPHVLARSALIAHSI-----LNSK 1807  
QY 988 QTVVCINI-YKIFLLQAYRFHACVI 1010  
Db 1808 HASLLNTRYISARKSFDYQKKIV 1831

RESULT 6  
EDBECA  
Immediate-early protein Iel175 - Marek's disease virus (strain GA)  
N: Alternate names: infected cell protein 4  
C: Species: Marek's disease virus  
C: Date: 30-Jun-1993 #sequence\_revision 30-Jun-1993 #text\_change 05-Sep-1997  
C: Accession: A42991  
R: Anderson, A.S.; Francesconi, A.; Morgan, R.W.  
A: Title: Complete nucleotide sequence of the Marek's disease virus ICP4 gene.  
A: Reference number: A42991; MUID: 92351564  
A: Accession: A42991  
A: Molecule type: DNA  
A: Residues: 1-1415 <AND>  
A: Cross-references: GB:M75729; NID:9330950; PID:9330951  
C: Superfamily: herpesvirus immediate-early protein Iel175  
C: Keywords: DNA binding; early protein; transcription regulation

Query Match 1.9%; Score 111.5; DB 1; Length 1415;  
Best Local Similarity 21.4%; Pred. No. 4.6; Mismatches 129; Indels 177; Gaps 25;  
Matches 97; Conservative 50;

QY 172 GSPLYQICATTDIWPVSASYSYRTPVGRNFTNLRLQIHKSSRQ-----EAPKP----- 223  
Db 52 GQFWFPLPVPD-WNPSTDWRSPP-----RSGPKKDFCGDLPAPLTSGP 96  
QY 223 -LALPSRGTKRHLSLTSTSVPSAKKARCYVPVPRVEEGPHRQVLTSPGSKSWPSPARSPE 281  
Db 97 RLTPSSGRMSLPHTTSS-----PRSSPRPR--GPE---TSPSNEHIIFSPRNP 143  
QY 282 VPTAEKDLSSKGVSDLSLGSVCCKHKPSSTSL-----SPPRQNA-FQL 326  
Db 144 SNTTHRVN---GHVSRSPSSSSSSSSSSSSSLVLSPPSSSRSPSPPPRADSS 200  
QY 327 RPFETRHFLYSRGDOERLNPFLSNLQPNLTGARRLVEIIFL-----GSRPRT 377  
Db 201 RP-----RRGRGSRGSGPQSGKGRKASPRTKLEDELYLQETANRRGGGRPG 251  
QY 378 SGPLCRTHRLSRR-----YWMRPL-----FOOL 401  
Db 252 RPP--KSGRAVQRNDIQVTSSSGLADTSPYDLGSGVMWEV-PLPFGPGRVFGGLGHRQA 308  
QY 402 LVNHAECQVRLRSHCHCRTANQVVT-----DALNTSPHMLDLRLHSSP 448  
Db 309 LTDSPE-----IVEAIHRFTNSHGVPVYVEEMKDYAKQYDALVNSLFH--KSKYNPLN 361  
QY 449 WQ-----VYGFRLACLKVVSAS-----LWGRTH- 473  
Db 362 WMHGLKSPADAALNHIYVQKFSQSYDSPAAGVGTIVNRCIPHAGAMKERKLLWAFPHI 421  
QY 473 -----NEERFEKLUKFI--SLGK-YGKLSIQE 497

```

QY 765 GQFLKHLQDSDASALRNSVVIQSIISMNESSSLDFELHFLHSVVKIGDRCVTQCGI 824
Db 1477 ----KLIHDESVS-----KLTYSSEASAA-----LHWVESTCAHIA-----TQCQYP 1514
QY 825 POGSSLSITLLCSLFCGDMENKLFVEYVRGDLGLLLRFDVDFLLVTPHLDOAKTFLSTLVHVG 884
Db 1515 AVGAKLQAIIA-----LVAKVK-----PKLESLSV-LENOQANNT 1547
QY 885 PEYGCMIINLQKTVNFPVEPQTL-GGAAPYQLPAHCLFPHWCGLLDLDQTLEVFCDYSGYA 943
Db 1548 DD-----INIIQAIS-SLDPRITGGTSKVQ-----EWTQVKDLLEK-----LA 1587
QY 944 QTSIKTSL--TFQSVFVAGTKMRN-KLLSVLRKLCGHLF----- 980
Db 1588 DTEASVQVNIQLRHMAITARSTNLNLANLKQKTSLEYKWKVKEHKTGPTSPITGIIKE 1647
QY 980 LDQVNSLOTVCINIIYKIFLQAYRFH-ACVIQLPDQVRKKNLTFELGIISQASCCYA 1038
Db 1648 LDLYLT-----FKLKEIYEYINQACIFS-----TFALPASKIDAKDALT 1687
QY 1039 ILKVKNGPMTLKASGFPP 1057
Db 1688 PLTPQSPVDFNTSGSPSP 1706

RESULT 8
A24938
hypothetical T2 protein - human herpesvirus 4
C:Species: human herpesvirus 4, Epstein-Barr virus
C>Date: 02-Jun-1988 #sequence_revision 02-Jun-1988 #text_change 29-Jan-1999
C:Accession: A24938; S33010; I49034
R:Bodescot, M.; Brison, O.; Perricaudet, M.
Nucleic Acids Res. 14, 2611-2620, 1986
A:Title: An Epstein-Barr virus transcription unit is at least 84 kilobases long.
A:Reference number: A24938; MUID:86176753
A:Accession: A24938
A:Molecule type: mRNA
A:Residues: 1-907 <BOD>
A:Cross-references: GB:X04060; NID:g59579; PID:g59580
A:Experimental source: strain B95-8
R:Farrell, P.J.
submitted to the EMBL Data Library, March 1988
A:Reference number: S32973
A:Accession: S33010
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-113 <FAR>
R:Cross-references: EMBL:V01555
R:Apolloni, A.; Moss, D.; Stumm, R.; Burrows, S.; Suhrbier, A.; Misko, I.; Schmidt, C.
Eur. J. Immunol. 22, 183-189, 1992
A:Title: Sequence variation of cytotoxic T cell epitopes in different isolates of Eps
A:Reference number: A49034; MUID:9211623
A:Contents: nuclear antigen EBNA-3
A:Accession: I49034
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 278-302 <APO>
A:Cross-references: GB:S7927; NID:g242887; PID:g242888
A:Experimental source: human A-type strain B95-8
A>Note: sequence extracted from NCBI backbone (NCBIN:79227, NCBI:P:79239)
C:Superfamily: human herpesvirus 4 nuclear antigen EBNA-3A

Query Match
Best Local Similarity 22.3%; Score 110; DB 2; Length 907;
Matches 82; Conservative 48; Mismatches 138; Indels 100; Gaps 20;

QY 38 LYQPGDPKPIYRILVAQCLVCMHWGSPPPADLSFQVSLKELVARVQVORCERNVL 97
Db 332 IVSRGPKFYKR-----PPI-----FIRLHRLMLMRAGKTEQ--- 365
QY 98 AFGFELLNEARG---GPPMAFTSSVRSYLPNTVITLTVRSWAMLLLSRVGDDLLVLLA 154

```

Mon Aug 16 10:21:27 1999

```

365 --GKEYLEKARSTGYTPRPVVKPRPEVPQ--DETATSHG--AAQPEPTIHLAA 416
155 HCALYLL-----VPPSCAYQVCGSPLYQICATTDIWPVSASYR-----PIRPVGRN 201
417 QGWAYPLHEQHGMACPVAQAAPTPLPV--SPGDQLPGVSDGRVACAPVAPAGPIVRP 475
202 FTNRLFOQIKSSRQEAEPKPLALPSRGTKRHLSLTSTSVPSAKKAR-CYVPVPRVEEGPH 260
476 W-----EPLTQAAGQAF--PVR--POHMPVEVPVPIVALERPVYKPK--VRPAPP 522
261 ROVLTPSGKS-----WVPSPARSEVPTAEDLSSKGKSDLSLSGV--- 305
523 KIAMOGGETSGIRARERWRPAPWTPNPRSP-----SQMSVRDLARLARAEACVQQA 576
305 CCKHKPSTSLSP-----PRNAFOLRPFIFTRHFLYSRGQERLNPSEL--LSN 354
577 SVEVQPQLTOVSPQPMEGPLVEQOMFPGAPFSQADV--RAGVPMQPVYFDLPL 634
355 LQNLTA 362
635 IQPISQA 642

RESULT 9
T00356
hypothetical protein KIAA0690 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 01-Feb-1999
C:Accession: T00356
R: Ishikawa, K.; Nagase, T.; Suyama, M.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.
A:Title: Prediction of the coding sequences of unidentified human genes. X. The complete
DNA Res. 5, 169-176, 1998
A:Reference number: Z14142
A:Accession: T00356
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1214 <ISH>
A:Cross-references: EMBL:AB014590; NID:d1204345; PID:d1032626
A:Experimental source: brain; clone HK03594
C:Genetics:
A:Note: KIAA0690

Query Match 1.9%; Score 109.5; DB 2; Length 1214;
Best Local Similarity 21.3%; Pred. No. 5.2;
Matches 121; Conservative 78; Mismatches 191; Indels 179; Gaps 31;

QY 23 PLATFVRRLGEGRRIVQPG---DPKIYRTLVAQCLVCM-HWGSQPPADLSFHOVS--- 76
DB 508 PLA---NTLKSAMDLAQAGSTVESKIYDTLOWQMTLLPGFCTRTDVAISFKGLARTL 564
QY 76 -----SLKELVARVQRLCERNE-----RNVLAFGFELINE--ARG--- 110
565 GMAISERPDRLVTVQALRTILITKGOAEADRAEVSFRKFNELPILFNLVYGOVVAAGDTP 624
QY 110 GPMNAFTSSVRSYLPNTVETLTVSGAWMLLSKVGDDLL-----VYLLAHCALVLYVP 163
625 APRRAVLETITYL--TITDTQLVNS-----LLEKASEKVIDPASPOTRVLVDLVLVA 678
QY 164 PSCAYQVCGSPLYQICATTDIWPVSASYRTPRVGRNFTNLFLQOIKSSRQEAAPKPL 223
679 P-CADAAIAISLY-----STIRPYLESKAGVQKKAIRVLEVCAS--PQPGAL 725
QY 224 ALPS--RGTKRHL--SLTSTVPSAK--KARC--YVPVPRVEEGPHQVLPITSGKSWVPSA 277
726 FVQSHLEDLKTLLDLSLSTSPAKRPLKCLLHIVRKLSEAHEFI-----TA 774
QY 278 RSEVPTAEKDLSSKGKVSDDLSSGVCCKKHKPSSTLSLSPRONAFOLRPFIFTRHFLY 337
775 LIPEVILCTKEVSGA-----RKNAPAL--LVEMGHAF 806
QY 338 SRGDGQERLNPFLSLNLOPNTGTARRLVEIFLGSRRPTSGPLCRTHLSRRYWMRPL 397

365 807 RFGNQEEALQCYLVL-TYPGLVGAVTMVS-----CSIALTHLFFEKGL 851
398 F-----QQLLVNHAECQVRLRLSRCHFRFTANQOVTDALNTSPPHLMDLLRLHSSPMQVY 452
852 MGTSTVEQULEN-----VCLLLASRTR-----DVKR-----SAL 880
453 GFLRACL-----CKVYSASLWGTNRHNERFF-----KNL-KKEISLGKYGKLSLQ 496
881 GFIKVAVTVMDVAHLAKHVOLVMEAGIKLSDDMRRHRMKLRNLTKFIR--KFG----F 934
497 ELMWKMKVEDCHWLRSRSPGKDRVPRAEHR 525
935 ELVKRLLPPEYHRVLYNIRKAEARAKHR 963

RESULT 10
S50754
hypothetical protein WP6 - Chlamydomonas eugametos
C:Species: Chlamydomonas eugametos
C:Date: 14-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 09-Sep-1997
C:Accession: S50754
R: Woessner, J.P.; Molendijk, A.J.; van Egmond, P.; Klis, F.M.; Goodenough, U.W.; Hari
Plant Mol. Biol. 26, 947-960, 1994
A:Title: Domain conservation in several volvoclean cell wall proteins.
A:Reference number: S50754; MUID:95093034
A:Accession: S50754
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-351 <WOE>
A:Cross-references: EMBL:L29028; NID:g530877; PID:g530878

Query Match 1.8%; Score 108.5; DB 2; Length 351;
Best Local Similarity 24.08; Pred. No. 0.97;
Matches 81; Conservative 28; Mismatches 144; Indels 85; Gaps 13;

QY 104 LNEARGGPMFAFTSSVRSYLPNTVI-ETLVRSGAWMLLSRVGDDLLVY-----LL 153
DB 26 LOGARAVPPSPAPACNLTTVNTAVNTQTLVNSAVLI--SVAPSYAVNASCLOSARGAL 83
QY 154 AHCALYLLVPPSCAYQVCGSPLYQ-----IC-----ATTDWP 186
84 GSCITDLGRTSGCCTQRCYTALHNLTLDCVNLVRAVCRMGSDGNATVNTTARAFTHPLF 143
187 SVSASYRTPRPVGRN---FTNRLFOQIKSSRQEAAPKPLALPSRGTKRHLSLTSTSVPS 243
144 SVYNNCIDTRPAPYNCSTENVTSVITPTSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 203
244 -AKKRCYFVPRVEEGPHRQVLPITPSGK-----SWVPSPARSEVPTA 285
204 PSPKASPSPKASAPSPAPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 262
286 EKDLSSKGKVSDDLSSGVCCKKHKPSSTLSLSP-----PRONAFOLRPFIFTRHFLYSRGDG 342
263 SPFSPASPSASP--SLSPKVSPTPTGTGSPAASPSGSPRAS-----PPGG 306
QY 343 QERLNPSELLNLOPNTGTARRLVEIFLGSRRPTSGP 380
DB 307 PPAMSP-----RUSPTPIPTSRPILPIPIPTPTSP 339

RESULT 11
S49627
regulatory protein ARGR2 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YML099c
C:Species: Saccharomyces cerevisiae
C:Date: 02-Dec-1994 #sequence_revision 10-Feb-1995 #text_change 12-Dec-1997
C:Accession: S49627; A25064
R: Gentles, S.; Bowman, S.
Submitted to the EMBL Data Library, November 1994
A:Reference number: S49627
A:Accession: S49627
A:Molecule type: DNA

```

A:Residues: 1-880 <GEN>  
 A:Cross-references: EMBL:Z46660; NID:9575702; PID:9575703; MIPS:YML099c  
 R:Messequy, F.; Dubois, E.; Descamps, F.  
 Eur. J. Biochem. 157, 77-81, 1986  
 A:Title: Nucleotide sequence of the ARGRII regulatory gene and amino acid sequence homol  
 A:Reference number: A25064; MUID:86220196  
 A:Accession: A25064  
 A:Molecule type: DNA  
 C:Residues: 1-3, 'F', 5-128, 130-131, 'L', 132-282, 'V', 284-344, 'V', 346-365, 'Q', 367-548, 'A', 55  
 C:Genetics:  
 A:Gene: SGD:ARG81; ARG82  
 A:Cross-references: MIPS:YML099c; SGD:S0004565  
 A:Map position: 13L  
 C:Superfamily: unassigned GAL4-type zinc cluster proteins; GAL4 zinc binuclear cluster h  
 C:Keywords: DNA binding; nucleus  
 F:16-53/Domain: GAL4 zinc binuclear cluster homology <GAL4>

Query Match 1.8%; Score 108; DB 2; Length 880;  
 Best Local Similarity 20.7%; Pred. No. 4.2;  
 Matches 98; Conservative 69; Mismatches 165; Indels 142; Gaps 26;

QY 93 ERNVLAFFGELLNEARGGPPMAFTSSVRSYLPNTVITLRSVGMMLLSRVGDDLLVYL 152  
 Db 469 QKFLGIGIELRQASNF--LRCLNTKSSIEPKYKDLTA-----ILSMNSIDVWGT 520  
 QY 153 LAHCALYLLVPPSCAYQVCGSPLYQIC-----ATTDIMPVSASVYRTPVGRNFTNLF 207  
 Db 521 MADCODHL-----ALCEDFVESRMKLRPNIS---EKTKLHRIEFLKL 561  
 QY 208 LQOIKSSROEAPKPLAPSRGTRKHLSL-TSTSPSARKARCYVPVVEGPHROVLEPT 266  
 Db 562 IDQSTALDVKRAKEIVLPSEEDNYPKLDTSNATTSSEPR---VDVVQEGFLREALNE 618  
 QY 267 PSKG---SWVPSPARSPEVPTAEKDLSSGKVDLSGVSCKKPKSPSTSLSPPRONA 323  
 Db 619 NGDKTHIEVKEP-----ITNVSAST-----PST--Tpp----- 648  
 QY 324 FOLRFETRHLYSRGQGERLNSFLLSNLQNLTGARRLV-----EIIFLGSR----- 375  
 Db 648 ---IFTNIATESY-YNKSD-----ISKLVKTDENIIGTDSYGLPNSLILFLSDCVRI 698  
 QY 375 -----PTSGPLCTHLSRRYQWRPLFQO-----LLVNHA 407  
 Db 699 RHNEYNYLTYPVKRFNELSINFELKLLKWKSEWNEHQENSEGKSFNSTAEALYHTM 758  
 QY 408 COYVRL-----RS-HCRF-RTANQOQVTDALNTSPHLMDLRLHSSP--WQVYGLR 456  
 Db 759 SFYSLIIYYFTMARSLNCQFLQNVVAKVLDHLN-AMEELVDQKKVIVPLIWO--GFMA 815  
 QY 457 ACLKVVASLWGTTHNERREFKNLKKFIS---LGKY-GKLSLOELMKMKVED 506  
 Db 816 GCAC-----TDENRQOEFRRAAKLAESGVGYWGARGVMLEWRRKKED 860

RESULT 12  
 T01397

LTR transposon polyprotein homolog T419.16 - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 24-Mar-1999  
 C:Accession: T01397  
 R:Parrell, L.D.; Gnoj, L.; de la Bastide, M.; Hameed, A.; Habermann, K.; Schutz, K.; Hua  
 submitted to the EMBL Data Library, May 1998  
 A:Description: Genomic sequence of BAC T419 from Arabidopsis thaliana, chromosome IV, ne  
 A:Reference number: Z14314  
 A:Accession: T01397  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1456 <PAR>  
 A:Cross-references: EMBL:AF069442; NID:93242970; PID:93924609  
 C:Experimental source: cultivar Columbia  
 C:Genetics:  
 A:Map position: 4

A:Introns: 129/1  
 A:Note: T419.16

Query Match 1.8%; Score 107; DB 2; Length 1456;  
 Best Local Similarity 19.7%; Pred. No. 11;  
 Matches 169; Conservative 108; Mismatches 288; Indels 294; Gaps 46;

QY 88 LCERNERNVLAFFGELLNEA---RGGPPMAFTSSVRSYLPNTV----- 128  
 Db 603 LSEKRRHIVEMGLTLLSHASVPKTYWYAF--SVAYILINRLPTLLOLQSPFQKLFQ 660  
 QY 128 ---JETLRVSCA---WMLLSR--VGDD-----LLVYLLAHCA-LYLLVPPSCAYQVCG 172  
 Db 661 PNEYKLVKFGCACYPLWRPNRHKLEDKQCAFMYSLTQSAYLCLHIPTGRLYTSRH 720  
 QY 173 SPLYQIC---ATTDIMPVSASVYR-----PTRP-----VGRNFTNLF 208  
 Db 721 VQDFERCFFSTTFNGVSTSQEQRSDSAPWPSHTTLPPLVLPAPPCGLPHL----- 775  
 QY 209 QOIKSSROEAPKPLAPLS-----RGTKRLSLTSTSVPSAKK--ARCYPVPRVEGPHR 261  
 Db 775 -----DTSPPRPPSPPLCTTOYSSSNLPSISSSPSSSEPTAPSHNGPQTAQPHQ 826  
 QY 262 -----QVLTPSGKSWVP-----SPARSPEVPTAEKDLSSGKVDLSL 302  
 Db 827 TQNSNSPILNPNPNSPSPNPNQSPPLQSPISPHIPTSTISE----- 876  
 QY 303 SVCKKHKPSTSLSPRONAFOLRPFETR-----HFLYSRG-DGOERLNPSF-LLS 353  
 Db 876 -----PNSPSSSTSTPLPVPAPPIQVNAQAPVNTSHMATRAKDGIRKPNQKYSAT 931  
 QY 354 NLQPNLTGARRLVEIFLGRSRTSGPLCTHLSRRYQWRPLFQOQLLVNHAECOYVRL 413  
 Db 932 SLAAN-----SEPTAIQAMKDDR-----WQAMGSEINQA----- 963  
 QY 414 LRSCHFRFTANOQVTDALNTSPH--LMDLLRLHSPWQVYGLFACLCCKVVSASLWCTR 471  
 Db 963 -----IGNHTWDLVPPPPSVTIIVGCRWIFTKKNSDGLNRYKARLVAG----- 1009  
 QY 472 HNERREFKNLKFISLGVKGLSL-----QELMKMKVEDCH-----WLRSSP 514  
 Db 1009 YNORGLDYAETFSPIVKSTSIKVLGVAVDRSWPIROLDVNNAFLOGTLTDEVIMSOP 1068  
 QY 515 G---KDRVFAAEHRLRERI-----LMTFLW-----LMDT--YVQVLLRSF 550  
 Db 1069 GFVDKDR-PDYVCLRLKAIYGLQAPRAYVELTYLLTVGFVNSISDTSFLVLRGSI 1127  
 QY 551 F-----YITESTFQKNLFFYKKSWSKLSQISGRQH-----LERVRLRE---LSQE- 595  
 Db 1128 IYMLYVDDILITGNDTVLLKHTLDALSQRFSVKEHEDLHVFLGIEAKRVPQGLHSORR 1187  
 QY 595 ---EYVRHHQDTWLAMPIC-----RLRF-----IPKNGLRPIV-NWSYSMTGR----- 634  
 Db 1188 YTLDLLARTNMLTAKPVATPMATSPKTLHSGTKLPDPTFYRGIVGSLQYLAFTRPDSLY 1247  
 QY 634 ALGRKQKQHQHTORLKTFLFSLNLYERTKHPHLMGSSVLGMNDIYRTWRAFLVRVRLDQ 693  
 Db 1248 AVNRLSQYMH-----MPTDD---HWNALKRVLRYLAGT 1277  
 QY 694 P-RMYFVKADVT---GAY-DAIPQGLVEVVAN-----MIRHSESTYCIQYAVVRRDS 742  
 Db 1278 PDGIFLKKGNLTSLHAYSADNAGDDDDYVSTNGYIVYLGHPHISWSSKKQGVVRSST 1337  
 QY 743 QGVHKSFRFRQVTTLSLQ 761  
 Db 1338 EAE-----YRSVANTSSSELQ 1352

RESULT 13  
 C71436

probable resistance gene - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse ear cress)

A.Variety: columbia  
C.Date: 03-Aug-1998 #sequence\_revision 03-Aug-1998 #text\_change 05-Dec-1998  
C.Accession: C71436  
P.Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirks, R.; Wedler, H.; Wedler, E.; Wambitt, R.; Weitzneger, T.; Pohl, T.; Terry, N.; Giel, A.; vanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.  
Nature 391, 485-488, 1998  
A.Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech, J.; Erhardt, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ans, C.; Chalwatzis, N.  
A.Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thaliana  
A.Reference number: A71400; MUID:98121113  
A.Accession: C71436  
A.Status: preliminary; nucleic acid sequence not shown; translation not shown  
A.Molecule type: DNA  
A.Residues: 1-1256 <BR>  
A.Cross-references: GB:297342; NID:g2245031; PID:e327519; PID:g2245043  
C.Genetics:  
A.Map position: 4COP9-4G3845

Query Match	1.8%;	Score 106.5;	DB 2;	Length 1256;
Best Local Similarity	19.0%;	Pred. NO. 9.3;		
Matches 216;	Conservative 154;	Mismatches 351;	Indels 417;	Gaps 57;

12	SLLSRYREVP	PLATVR	REGL	VEG	RL	VQ	GP	KI	YR	TL	-----	VAOCLVM	58																																									
388	SSUKGRD	KDW	-----	KMP	---	RL	NSD	DK	IE	TL	RV	GR	LK	NK	RE	LK	FI	CI	AC	438																																		
59	HWSQ	PP	AD	LS	PH	QV	SS	LK	EL	VAR	W	---	Q	RL	C	E	R	N	E	R	N	V	L	A	G	---	F	E	L	L	---	106																						
439	FNG	-----	---	F	K	V	S	N	K	E	L	L	---	D	V	G	L	T	---	P	D	G	I	E	H	M	L	L	E	K	L	G	R	E	D	488																		
106	-	E	A	R	G	P	P	M	A	F	T	S	S	V	R	S	V	L	P	N	T	I	E	T	L	R	V	S	G	A	M	L	L	S	R	V	G	D	L	L	Y	L	L	A	H	C	A	L	Y	L	L	P	163	
489	RAK	S	G	N	P	-----	A	K	R	Q	F	L	N	---	F	E	D	I	Q	---	E	V	V	T	E	K	T	G	T	V	---	L	G	I	R	V	P	527																
164	P	S	C	A	Y	Q	V	G	S	P	L	Q	I	C	A	T	T	D	I	M	P	S	V	S	A	S	V	P	R	P	V	---	G	R	N	E	T	L	R	F	L	O	O	I	S	S	S	R	Q	E	A	219		
528	P	T	V	L	F	S	-----	T	R	P	L	V	I	N	E	S	F	G	M	R	N	L	O	Y	L	E	I	G	H	W	S	E	I	562																				
220	P	K	L	A	L	P	S	R	G	T	K	R	H	L	S	T	---	T	S	V	P	S	A	K	A	---	-	R	C	P	V	P	R	V	E	B	G	P	H	O	V	L	P	T	P	267								
563	D	L	P	Q	L	V	V	L	P	L	K	L	K	W	N	C	P	L	S	P	F	T	K	A	E	Y	L	N	I	M	K	Y	S	K	L	E	K	L	W	E	---	T	L	P	L	G	618							
268	S	K	S	W	---	V	P	S	P	A	R	S	E	V	P	T	A	E	K	L	S	S	G	K	Y	S	D	L	S	G	S	C	C	K	H	K	S	T	S	L	S	P	P	P	R	Q	N	A	F	Q	L	326		
619	S	L	K	M	D	G	C	S	N	N	K	E	I	P	---	D	L	S	L	A	I	N	E	E	L	N	S	K	E	S	L	T	P	S	S	---	O	N	A	I	K	L	667											
327	R	P	F	I	E	T	R	F	L	X	S	R	G	D	G	E	R	L	N	P	S	F	L	S	N	O	P	N	T	G	A	R	L	V	E	I	---	I	F	L	G	S	R	P	T	S	G	P	L	C	R	T	H	385
668	R	T	-----	L	C	S	G	---	V	L	I	D	L	K	---	S	L	E	G	M	C	N	E	I	L	S	V	D	W	S	N	E	T	G	L	I	V	P	R	710														
386	R	L	S	R	E	Y	Q	M	R	P	L	F	O	Q	L	N	H	A	E	Q	Y	V	R	L	R	S	H	C	R	F	T	A	N	Q	O	T	D	A	L	N	S	P	H	M	D	L	R	L	H	445				
711	K	L	K	R	L	W	D	Y	C	P	---	K	R	L	S	N	E	F	A	E	Y	L	V	E	L	R	M	E	---	N	S	D	L	E	K	L	W	---	G	T	P	L	G	S	L	K	E	M	Y	L	H	763		
446	S	S	P	W	-----	---	Q	V	G	F	I	R	A	C	L	K	V	S	A	S	L	W	T	R	---	H	N	E	R	R	E	F	K	N	L	K	F	I	S	L	G	K	485											
764	G	S	K	Y	L	K	E	I	P	D	L	S	A	I	N	E	R	L	Y	F	O	C	E	S	L	T	P	S	I	O	N	A	T	L	N	D	R	O	C	K	L													

QY 634 ALGRKQAQHFTQRLKTLTSLMLNYETKHPHLMGSSVLGMNDIYRTWRAFLVLRALDQT 693  
Db 962 HRLVRLMEKCT-GLELLPTDVLNLSLLILDLSGCSSL-----RTFPLISTRIECL--- 1012  
QY 694 PRMYFVKADVTGAYDAIPQGLKEVYVANMRHSESTYCIQYAVVRRSQGVHKSFRQ 753  
Db 1012 --YLENT---AIEEVP-----CCI-----ED 1027  
QY 754 VTTLSLDLPYMGQFLKHQLQSDASALRNSVWIEQISIMNESSSLDFDLFLHRSVVKI 813  
Db 1028 LTRUSVLMLCCQRLKNI-----SPNIF-----RLTSLMV 1057  
QY 814 GDRGYTCOGIPQGSLSLTLLCSLGFDMENKLEFAEQVRDGLLLREVDDELVTPLHDAQ 873  
Db 1058 AD--FTDCRGVIKALSDATVATM-----EHVSCVP----- 1088  
QY 874 KTFSLTVHGVPEYGC-----MINLQKTVVNPVPEPTLGGAPYQLPAHCLF----- 922  
Db 1088 ---LSENI---EYTCERFWDELVERNSRISFYKDEGDG-----PEDLPKSLTFNDVEF 1135  
QY 922 -----PWGGLLDQTQLEVFDYSGYAQTSJK-----TSLTF-----QSVFKAG 960  
Db 1136 KFCSSNRRIKECGVRLLYVVSQET--EYNQOTRSKKRMRTSGTSEYINLAGDOIADTG 1193  
QY 961 KTMRNKLLSV-----LRLKCHGLFLDLQVNSLQVTCINIKYIFLLQAYRHACVI 1010  
Db 1194 LAALNLELSLGGEPSSTSLEGEALCYDVYMITEDQKCI-----LFLVTEASYHSFLV 1247  
  
RESULT 14  
A45761  
Ca2+-transporting ATPase (EC 3.6.1.38) - Plasmodium yoelii  
C:Species: Plasmodium yoelii  
C:Date: 03-Jun-1993 #sequence\_revision 03-Jun-1993 #text\_change 21-Aug-1998  
C:Accession: A45761  
R:Murakami, K.; Tanabe, K.; Takada, S.  
J. Cell Sci. 97, 487-495, 1990  
A:title: Structure of a Plasmodium yoelii gene-encoded protein homologous to the Ca(2)  
A:Reference number: A45761  
A:Accession: A45761  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1115 <MUR>  
A:Cross-references: GB:X55197; NID:g10097; PID:g10098  
C:Superfamily: Na+/K+-transporting ATPase alpha chain; ATPase nucleotide-binding domain  
C:Keywords: hydrolase  
S:689-866/Domain: ATPase nucleotide-binding domain homology <ATP>

Query Match 1.8%; Score 106; DB 2; Length 1115;  
Best Local Similarity 19.3%; Pred. No. 8.5;  
Matches 11; Conservative 87; Mismatches 192; Indels 186; Gaps 30;

Qy	556	STQKNRL-----PFYRKSVMWSKLSIGVQHLEHRLVRLELSQEEVRRHHODTWL-----	603
Db	554	STYKKNISDKSEPTFPKCVSAWNECTIMRIIEFTRERKLMSVWVENSNEYLYCK	613
Qy	605	AMP-----ICRLRFPKPNGLRP-----IVNWSYSMGTRALGRRKQAQHFOTURUKTLFS	653
Db	614	GAPENIINRCKYMSKNDIRPLDLSKLNELNKINMGKRA-----LRTL-----	659
Qy	654	MLNVERTKHPHLMGSSVGLMNDIYRTWFAFV-LRVRALDQTPRMVEVKA-----	702
Db	659	SFAYKKYKNSDI---NIKNSEDIYKLEHDLITYIGGLIIDPFRKYVGKALSCHLAGIRV	715
Qy	702	-DVTGAYDAIPOQKLVEVAVNMRHSES-----TYCI---ROYAVRRDSQCVHKSPROVT	755
Db	716	FMITG--DNIDTAKAIKEINILNHDDTDKYSCFNGREFFEDLPEBKQYILNKY---QQIV	772
Qy	756	TLSDLDQYMGQFKHLQD-----SDASALRNS-VWIEQSI-SMNNESSLSLDFEF	802
Db	773	FCRTEPKHKNIVKILDLGETVAMTGDGVNDAPAKSADIGATMGINGTAVKAEASD--	831

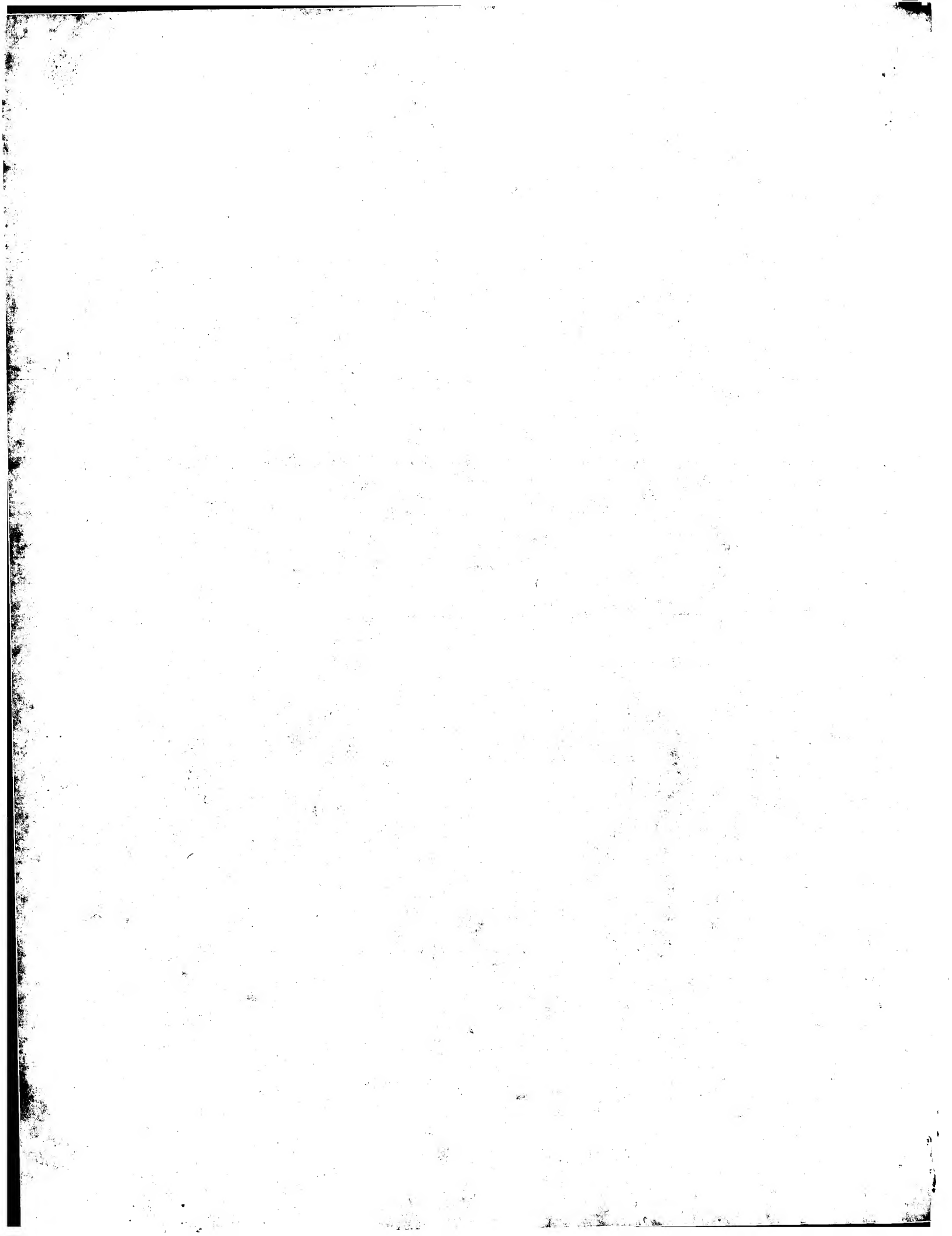
QY 803 LHLRHSVVKIGRCYTOCQIQPOGSSLSLTLCSLCFGDMENKLFPAEVQRDGLLLREVDD 862  
Db 831 -----LILADDNFTIVEAIKEGRCI-----YNNMK-----AFIRY--- 862  
QY 863 FLLVTPHLDQAQTLSTLVHGVPEYGCMLNQKTVWNPVE--PGLTGGAPYQLPAHCL 920  
Db 862 --LISSNIGEVASIFITAILGIPD--SLAPVOLLVNLVTDGLPATALGNP---PEHDV 914  
QY 921 FPMC-----GILL-----DTOTLEVFCDYSG 941  
Db 915 MK-CKPRHRNDNLINGLTLRYIVIGTVGIATVSIYIYWMFYDPMDNHTLINFYQLSH 973  
QY 942 YAOISIKTSLTFQ-----SVEKAGKTMKNKL-LSVLRL-----KCHG 977  
Db 974 YNQ--CKTWSFNVNKYVDMSEDLCYSFSAGKVKASTLSLSVLVIEMFNALNALSEYNS 1031  
QY 978 LFLDLQVNSLQTVCIINYKIFLQOAYRFHACVQLP 1013  
Db 1032 LFLVPPWRNMYLVLATIGSLFL-----HCLIIYFP 1061

## RESULT 15

S40013  
hypothetical protein 2 - Calothrix sp. (PCC 7601)  
C:Species: Calothrix sp.  
A:Variety: PCC 7601  
C:Date: 08-Jun-1994 #sequence\_revision 12-Apr-1996 #text\_change 17-Mar-1999  
C:Accession: S40013; S35079  
R:Ferat, J.  
submitted to the EMBL Data Library, April 1993  
A:Reference number: S40013  
A:Accession: S40013  
A:Molecule type: DNA  
A:Residues: 1-584 <FE1>  
A:Cross-references: EMBL:X71404; NID:g396226; PID:g396228  
A:Experimental source: PCC 7601  
R:Ferat, J.L.; Michel, F.  
Nature 364, 358-361, 1993  
A:Title: Group II self-splicing introns in bacteria.  
A:Reference number: S35078; MUID:933233980  
A:Accession: S35079  
A:Molecule type: DNA  
A:Residues: 224-338:513-565 <FE2>  
A:Cross-references: EMBL:X71404  
A:Experimental source: PCC 7601

Query Match 1.8%; Score 106; DB 2; Length 584;  
Best Local Similarity 22.5%, Pred. No. 3.2;  
Matches 62; Conservative 33; Mismatches 111; Indels 70; Gaps 11;  
QY 676 YRTWRVFLRVRLDQTPRMFYKADVTGAYDAIPQGLVEYVANMIRHSESTYCIROY 735  
Db 50 LMRWSGKALAVRKVTO-DNQGKKAAGIDGVKSLKPSARLTLVNMKNLHK-----VKAT 103  
QY 736 AVVRDSOGVHKSFRRQVTTLSL-----LQPYMGQFLKHLQSDA- 777  
Db 104 RRVWPKFGNVEKR-PLGIPTMQDRATQSLVKALEPEWEAKFEPNSYGFPRGNADAR 162  
QY 777 SALRNSVVEQISISMNESSSLDFDFLH-----FLRHSVVKIGDRC 817  
Db 163 EAFINSIRYSNKKVLDADISKCFKINHEKLLTINTFTPMRROIKAWLKAGVLDNGHFS 222  
QY 818 YTCQGGIPOGSSLSLTLCSLCFGDMEN--KLFAEVQRD-----LLRFVDDFLVLT 867  
Db 223 ET-TEGTPQGGVISPILLANIALHGLEKLVKEFAASQGGKVKQNQNSISLIRYADDEVILA 281  
QY 868 PHLDQ-----AKTFLSTLVHGVPEYGCMLNLOKT 896  
Db 282 PNKTQIIIVLKEIVKTLA-----EMGLEPNKNT 310





GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: August 14, 1999, 04:51:16 ; Search time 12.16 Seconds  
(without alignments)  
2608.308 Million cell updates/sec

Title: US-09-042-460-2  
Perfect score: 5901  
Sequence: 1 MTRAPRCPAYRSLRSRYRE.....TILKAAADPALSTDFQILD 1122  
Scoring table: BLOSUN62

Searched: 77977 seqs, 28268293 residues  
Database : SwissProt 37.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query No.	Score	%		Length	DB	ID	Description
			Match					
1	2	124	2.1	4540	1	DYHC_PARTE	Q27171	paramecium
2	1	115	1.9	1415	1	ICP4_EBVMG	Q02362	marek's dis
3	3	110	1.9	812	1	EBN3_HSV	P12977	epstein-bar
4	4	108	1.8	880	1	ARG2_YEAST	P05085	saccharomyc
5	105	105	1.8	1493	1	MEK1_RAT	Q22935	rattus norv
6	6	104.5	1.8	916	1	RTJK_PROFU	P13329	drosophila
7	103.5	103.5	1.8	719	1	DEND_RAT	P30617	rattus norv
8	102.5	102.5	1.7	758	1	DEND_RAT	P32622	neurospora
9	101	101	1.7	677	1	CY14_NEUCR	P49848	homo sapien
10	99.5	99.5	1.7	810	1	T2D5_HUMAN	P15274	saccharomyc
11	98.5	98.5	1.7	2163	1	ADMS_YEAST	P32639	saccharomyc
12	98	98	1.7	745	1	BRR2_YEAST	Q13617	homo sapien
13	98	98	1.7	1239	1	CUL2_HUMAN	Q01098	mus musculu
14	97.5	97.5	1.7	916	1	NME3_MOUSE	P21328	drosophila
15	97	97	1.6	3148	1	RTJK_DRONE	P51112	fugu rubrip
16	97	97	1.6	174	1	HD_FUGRU	P18294	strongyloce
17	97	97	1.6	1314	1	KCRF_STRPU	P25390	saccharomyc
18	96.5	96.5	1.6	585	1	SS22_YEAST	P49092	lotus japon
19	96.5	96.5	1.6	446	1	ASN1_LOTJA	P09653	gallus gall
20	96.5	96.5	1.6	856	1	TBB5_CHICK	Q23023	caenorhabdi
21	96.5	96.5	1.6	1336	1	UN51_CAEEL	P53767	rattus norv
22	96.5	96.5	1.6	2870	1	VEGR_RAT	Q10105	schizosacch
23	96	96	1.6	480	1	YAOS_SCHPO	P79892	equus cabal
24	96	96	1.6	1155	1	P53_HORSE	P53046	saccharomyc
25	96	96	1.6	2156	1	R0M1_YEAST	P27176	puumala vir
26	96	96	1.6	320	1	RRPL_PUUMH	P23070	escherichia
27	96	96	1.6	976	1	RT86_ECOLI	Q15431	homo sapien
28	96	96	1.6	865	1	SOPI_HUMAN	P06612	escherichia
29	96	96	1.6	1277	1	TOPL_ECOLI	Q04264	saccharomyc
30	95.5	95.5	1.6	714	1	YMW6_YEAST	P47597	mycoplasma
31	95.5	95.5	1.6	606	1	CLPB_MYCOE	P10398	homo sapien
32	95.5	95.5	1.6	1233	1	KRAA_HUMAN	P13368	drosophila
33	95	95	1.6	2554	1	NME3_HUMAN	P55198	homo sapien
34	95	95	1.6	1093	1	7LES_DRONE	P20951	papaya mosa
35	95	95	1.6	1547	1	AF17_HUMAN	P08548	nycticebus
36	94.5	94.5	1.6	1260	1	RRO_PMW	Q94637	oschelus br
37	94.5	94.5	1.6	1660	1	LINI_NYCCO	P40456	saccharomyc
38	94.5	94.5	1.6	1118	1	VIT6_OSCBR	P15822	homo sapien
39	94.5	94.5	1.6	2717	1	Y1P1_YEAST	P42684	homo sapien
40	94	94	1.6	1182	1	ZEP1_HUMAN	P35658	homo sapien
41	94	94	1.6	2090	1	ABL2_HUMAN	P08487	bos taurus
42	94	94	1.6	1291	1	N214_HUMAN	P12689	saccharomyc
43	94	94	1.6	985	1	PIP4_BOVIN		

44	94	1.6	445	1	TBB_PLAFK	P14643	plasmodium
45	94	1.6	656	1	UL25_HCMVA	P16761	human cvtom

## ALIGNMENTS

```

RESULT 1
DYHC_PARTE
ID DYHC_PARTE STANDARD; PRT; 4540 AA.
AC Q27171;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35; LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36; LAST ANNOTATION UPDATE)
DE DYNEIN HEAVY CHAIN, CYTOSOLIC (DYHC).
DS DHC-8.
OS PARAMECIUM TETRAURELIA.
SU EUKARYOTA; ALVEOLATA; CILIOPHORA; NASSOPHOREA; PARAMECIUM.
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN-STOCK 51;
RC MEDLINE: 96157890.
RT KANDL K.A., FORNEY J.D., ASAI D.J.;
RA "The dynein genes of Paramecium tetraurelia: the structure and
RT expression of the ciliary beta and cytoplasmic heavy chains.";
RL MOL. BIOL. CELL 6:1549-1562(1995).
CC -!- FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES.
CC -!- SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -!- SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).

```

Query Match 2.1%; Score 124; DB 1; Length 4540;

Matches	127;	Conservative	111;	Mismatches	214;	Indels	234;	Gaps	33;
---------	------	--------------	------	------------	------	--------	------	------	-----

[illegible]

QY 542 YVQLLRSEFYTES-----TFQK---NRLFFVRK-----SVWSKLOSIGVROHLE 584  
 Db 570 YOSQLTKTVHKDIOSLQNKFEKTYOKSONSRASADIPLTSGFVWSKQLOIRLOKQWQ 629  
 QY 585 RV-----RLRELSQ-----EEVRHHDOTWLAMPICRLRF 613  
 Db 630 KVEQILQPOWAEYDQCKCKEMGEFETILDSGPALEDWKQEIHNHN-----KA 678  
 QY 614 IPKPNGLRPIVNMYSMGTRALGRRKQAOHQHTORLKLFSMLNRYETKPHLMGSSVLGM 673  
 Db 679 VSONEKLFEVTVRRGLEIRVNYEKKLSQLE-KEVRNLSNM-----KTKVPYSISHIA--- 731  
 QY 674 NDIYRTWRAFLVRALDQTPRMFYFKADVTGAYDAIPOQKLVVEVYVANNIRHSESTYICIR 733  
 Db 731 NDAKAST-PFAL---SLQESLHYI---QTSOLNA-KSAKLV----- 766  
 QY 734 QYAVVRDSQGVHKSFRROVTTLSLOPYMGQFLKHL----- 772  
 Db 766 --AALRKEVOLOIGOGFNYLWHTKTOLOPYVKFTDKVFELEQAVNGLNERIGQIESLCE 823  
 QY 772 -----QSDASALRNSVIEQISIMNESSLSLDFELHFLRHSVVKIGDRCYTQCOGIP 825  
 Db 824 AMKTCPVDSLADKLKQIEVIDSLCFNNFSN-----LHWIOTDIDK----- 865  
 QY 826 QGSSLSLTLCSLCFGDMENKL-----FAEQVORDGLL-----LRFVDFFLLVTPHLDQA 873  
 Db 865 ---QIESILCDRVTVQMKELWQNFQVQKIQERGLVQNTVVHKLQDQIIYVDPPEYA 921  
 QY 874 KTFSLTVHGVPEYGCMMINLOKTVVN 899  
 Db 922 KYFWQEFHKM--IGQICSLPRLVAN 945  
 RESULT 2  
 ICP4\_HSVMG STANDARD; PRT; 1415 AA.  
 AC Q02362;  
 DT 01-OCT-1993 (REL. 27, CREATED)  
 DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)  
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)  
 DE TRANS-ACTING TRANSCRIPTIONAL ACTIVATOR PROTEIN ICP4 (IMMEDIATE-EARLY  
 DE PROTEIN IE1/5).  
 GN ICP4.  
 OS MAREK'S DISEASE HERPESVIRUS (STRAIN GA) (MDHV).  
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;  
 OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE; 92351584.  
 RA ANDERSON A.S.; FRANCESCONI A.; MORGAN R.W.;  
 RT "Complete nucleotide sequence of the Marek's disease virus ICP4  
 gene.";  
 RL VIROLOGY 189:657-667(1992).  
 CC -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE  
 CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING  
 CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS. IT IS  
 CC REQUIRED FOR THE SWITCH FROM IMMEDIATE-EARLY TO EARLY MODE OF  
 CC GENE EXPRESSION.  
 CC -!- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.  
 CC -!- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF  
 CC PHOSPHORYLATION.  
 CC -!- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY OF PROTEINS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M75729; G330951;  
 DR PIR; A42991; EDBEGA.

KW EARLY PROTEIN; TRANSCRIPTION REGULATION; TRANS-ACTING FACTOR;  
 KW DNA-BINDING; PHOSPHORYLATION; NUCLEAR PROTEIN.  
 FT DOMAIN 155 200 SER/PRO-RICH.  
 SQ SEQUENCE 1415 AA; 154936 MW; 21FB00A6 CRC32;  
 Query Match 1.9%; Score 111.5; DB 1; Length 1415;  
 Best Local Similarity 21.4%; Pred. No. 3.6; Mismatches 129; Indels 177; Gaps 25;  
 Matches 97; Conservative 50;  
 QY 172 GSPLYQICAITDIWPSVASYRTPRPVGRNFTNLRFQIQKSSSRQ-----EAPKP----- 223  
 Db 52 GQMPFPLPVD--WNSTDMRPSPP-----RSGPKKDFCGDLAPLTS GP 96  
 QY 223 -LALPSRGTKRHLSTSTSVPSAKKARCYVPVRYEGPHROVLPTPSGKSWPSPARSPE 281  
 Db 97 RLTPSSGRMSELPHTTSS-----PRSSRPRL--GPE---TSPNEHIITSPRNP 143  
 QY 282 VPTAEKDLSSKGYS-DLSLSGVCCXKHKPSSTSL-----SPPRONA-FQL 326  
 Db 144 SMTTRNV---GHVSRSPSS 200  
 QY 327 RPFETIRHFLYSGDQERLNPSELLSNLQNLGARRLVEIIF-----GSRPT 377  
 Db 201 RP-----RRGGSNRRGSGSPQSKGRKASPRTRKLEDEYLPQETANRRGGRRPG 251  
 QY 378 SGPLCRTHRLSR-----YQMRPL-----FQOL 401  
 Db 252 RPP--KSGRAVQRNDIQVTSSGLADTSPYDLGGSVMWEV-PLPPPGRCWFGGLGGRQA 308  
 QY 402 LYNHAEQYVRLLRSHCRFTANQOVT-----DALNTSPHLLDLRLHSSP 448  
 Db 309 LTDSPE-----IVEAIHRFTSHGPPVYVEEMKDYAKQYDALVNSLFH--KSMKVNPLN 361  
 QY 449 WQ-----VYGFRLACLCCKVVSAS-----LWGTRH- 473  
 Db 362 WMHGGKLSPADAAALNHVYQKFSYSDSPGAATGTVNRCPHITAGAMKERKLLWAFPHI 421  
 QY 473 -----NERFFKNLKKFI--SLGK-YGKLSLOE 497  
 Db 422 AASIATMTRYCKDQKTFRLRSUKKAYASWAFPD 454  
 RESULT 3  
 EBN3\_EBV STANDARD; PRT; 812 AA.  
 AC P12977; Q66540;  
 DT 01-OCT-1989 (REL. 12, CREATED)  
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)  
 DE EBNA-3 NUCLEAR PROTEIN (EBNA-3A).  
 GN BERTI.  
 OS EPSTEIN-BARR VIRUS (STRAIN B95-8) (HUMAN HERPESVIRUS 4).  
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;  
 OC GAMMAHERPESVIRINAE; LYMPHOCRYPTOVIRUS.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE; 84270667.  
 RA BAER R.; BANKIER A.T.; BIGGIN M.D.; DEININGER P.L.; FARRELL P.J.;  
 RA GIBSON T.J.; HATFULL G.; HUDSON G.S.; SATCHWELL S.C.; SEGUIN C.;  
 RA TUFFNELL P.S.; BARRELL B.G.;  
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";  
 RL NATURE 310:207-211(1984).  
 RN [2]  
 RN CHARACTERIZATION.  
 RX MEDLINE; 86287322.  
 RA HENNESSY K.; WANG F.; BUSHMAN E.W.; KIEFF E.;  
 RT "Definitive identification of a member of the Epstein-Barr virus  
 RT nuclear protein 3 family.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:5693-5697(1986).  
 RN [3]  
 RN SUBCELLULAR LOCATION.  
 RX MEDLINE; 90266473.

```

RA PETTI L., SAMPLE C., KIEFF E.;
RT "Subnuclear localization and phosphorylation of Epstein-Barr virus
AT latent infection nuclear proteins.";
RL VROLOGY 176:563-574(1990).
CC -I- FUNCTION: INVOLVED IN LATENT CYCLE.
CC -I- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH THE NUCLEAR MATRIX.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this Statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: V01555; E25053; ALT_INIT.
KW NUCLEAR PROTEIN.
SQ SEQUENCE 812 AA; 89097 MW; B8EFBDC6 CRC32;

```

[2]

RN SEQUENCE FROM N.A.  
RC STRAIN-S288C / AB972;  
RA GENTLES S., BOWMAN S., BARRELL B.G., RAJANDREAM M.A.;  
RL SUBMITTED (NOV-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.  
CC -!- FUNCTION: THIS PROTEIN REPRESS OR INDUCES SEVERAL GENES OR  
CC ARGININE METABOLISM.  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
CC -!- SIMILARITY: CONTAINS A ZN(2)-CYS(6), FUNGAL-TYPE BINUCLEAR  
CC CLUSTER DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X03940; G3378; -;  
DR EMBL; Z46860; G575703; -;  
DR PIR; A25064; A25064.  
DR SGD; L0000114; ARG82.  
DR PROSITE; PS00463; ZN2\_Cy6\_FUNGAL\_1; 1.  
DR PROSITE; PS50048; ZN2\_Cy6\_FUNGAL\_2; 1.  
DR PEAR; PF00172; Zn\_clus; 1.  
DR TRANSFAC; T00044; -;  
KW TRANSCRIPTION REGULATION; ACTIVATOR; DNA-BINDING; NUCLEAR PROTEIN;  
KW ZINC; METAL-BINDING; ARGININE METABOLISM.  
FT DNA\_BIND 21 48 ZN(2)-CYS(6), FUNGAL-TYPE.  
FT SIMILAR 587 690 TO N.CRASSA NMR (AA 96-193).  
FT CONFLICT 4 4 S -> F (IN REF. 1).  
FT CONFLICT 129 129 MISSING (IN REF. 1).  
FT CONFLICT 283 283 F -> V (IN REF. 1).  
FT CONFLICT 345 345 D -> V (IN REF. 1).  
FT CONFLICT 366 366 E -> Q (IN REF. 1).  
FT CONFLICT 549 549 T -> A (IN REF. 1).  
FT CONFLICT 597 597 T -> S (IN REF. 1).  
FT CONFLICT 665 665 K -> N (IN REF. 1).  
FT CONFLICT 869 869 V -> I (IN REF. 1).  
SQ SEQUENCE 880 AA; 100281 MW; CDBA1C8D CRG32;

RESULT	4
ARG2_YEAST	
ID	ARG2_YEAST
AC	PO5085;
DT	01-JAN-1988 (REL. 06, CREATED)
DT	01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
DE	15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE	DE ARGININE METABOLISM REGULATION PROTEIN II.
GN	ARGR2 OR ARG81 OR YML099C.
OS	SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
QC	EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
QC	SACCHAROMYCETACEAE; SACCHAROMYCES.
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE; 86220196.
RA	MESSENGUY F., DUBOIS E., DESCAMPS F.:
RT	"Nucleotide sequence of the ARGRII regulatory gene and amino acid
RT	sequence homologues between ARGRII PPRI and GAL4 regulatory
RL	proteins.";
RL	EUR. J. BIOCHEM. 157:77-81(1986)

Mon Aug 16 10:21:27 1999

Db 759 SFYSLIIYFTMARSLNCOFLQNVAKVLDHLN-AMEELVDQKKVIVPLIQ--GFMA 815

Qy 457 ACLCKVWSASLWGTNRHNRFFKLNLFKIFIS---LGKY-GKLSQELMMKMKVED 506

Db 816 GCAC-----TDENRQEFRRWAALAESGVSYNGARQVLMLEVRRRKED 860

```

RESULT 5
MEK1_RAT
ID MEK1_RAT STANDARD; PRT; 1493 AA.
AC Q62925;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE MAPK/ERK KINASE 1 (EC 2.7.1.1) (MEK 1)
GN MEK1 OR MEK2.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RC TISSUE=BRAIN;
RX MEDLINE; 96224276.
RA VANDERBILT C.A., COBB M.H.;
RT "Cloning of rat MEK kinase 1 cDNA reveals an endogenous membrane-associated 195-kDa protein with a large regulatory domain.";
RL PROC. NATL. ACAD. SCI. U.S.A. 93:5291-5295(1996)
CC -!- FUNCTION: CAN PHOSPHORYLATE AND ACTIVATE MAPKK 1 AND MAPKK 2 (MEK1/MEK2) WHICH LEADS TO PHOSPHORYLATION OF MAP KINASES. MOST POTENTIALLY ACTIVATES THE JNK/SAPK GROUP OF MAP KINASES, AND LESS EFFICIENTLY ERK2 OR P38.
CC -!- SUBCELLULAR LOCATION: MEMBRANE ASSOCIATED.
CC -!- TISSUE SPECIFICITY: MOST HIGHLY EXPRESSED IN SPLEEN, KIDNEY, AND LUNG.
CC -!- PTM: AUTOPHOSPHORYLATED.
CC -!- SIMILARITY: WITH THE CONSERVED CATALYTIC DOMAINS OF SER/THR-PROTEIN KINASES. BELONGS TO THE MAP KINASE KINASE FAMILY.
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U48596; G1354137; -
CC DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
CC DR PFAM; PF00069; Rkinase; 1.
CC DR HSPF; P24941; 1AQL.
CC KW TRANSFERASE; SERINE/THREONINE-PROTEIN KINASE; ATP-BINDING; PHOSPHORYLATION.
CC FT DOMAIN 2 5
CC FT DOMAIN 17 20 POLY-ALA.
CC FT DOMAIN 25 29 POLY-GLY.
CC FT DOMAIN 74 149 PRO-RICH.
CC FT DOMAIN 111 114 POLY-PRO.
CC FT DOMAIN 233 291 PRO-RICH.
CC FT DOMAIN 412 421 POLY-SER.
CC FT DOMAIN 439 555 PH.
CC FT DOMAIN 643 750 PH.
CC FT DOMAIN 1163 1168 POLY-GLU.
CC FT DOMAIN 1197 1200 POLY-ILE.
CC FT DOMAIN 1224 1489 PROTEIN KINASE.
CC FT DOMAIN 1230 1237 ATP (BY SIMILARITY).
CC FT BINDING 1253 1253 ATP (BY SIMILARITY).
CC FT ACT_SITE 1350 1350 BY SIMILARITY.
CC FT MUTAGEN 1369 1369 D->A; INACTIVATION.
CC SEQUENCE 1493 AA; 161315 MW; 982785D9 CRC32;

```

```

Query Match 1.8%; Score 105; DB 1; Length 1493;
Best local similarity 18.5%; Pred. No. 12;
Matches 114; Conservative 79; Mismatches 209; Indels 214; Gaps 30;

QY 163 PPSYAYVCGSPLYQICATTDIWPVSASIRP---TRPVGRNFTNLRLEQOI-KSSSROE 218
DB 112 PPPGAASRCGSHSAELAAARDGASPAAGAEPPSAAPSGREMENKETLGLHKMDRPE 171
QY 219 APKPLALPSRGTKRHLSTSTSVPSAKKARCPVPVPE----- 257
DB 172 -----ERMIREKLKATCMPANKHEWLERENRRGPVVVPIPIKGDG 212
QY 257 -----EGPHRQVLPTPSGKSWPSPARSPEVPTAEKDLSS---GKVSQSLSL-S 301
DB 213 SEMSNLAELQEGGAGSAAPAKGRR-SPSPGSSPSGRSGKPESGVRKRKRVPVFFOS 271
QY 302 GSVCCCKHKPSTSLSP--PRONAFOLRPFIETRHLYSRGQGOERLNP-SFLLSNLQPN 358
DB 272 GRITPPRRAPSPDGFSYSPETSRRVNVKVRARLYLL-----QQIGPNSFLIGDSDP 325
QY 359 LTGARRLVEIIFLGSRPRTSGPLCRTHRLSRRYWQMRPLFQQLLVNHAECQYVRLLSHC 418
DB 326 --NKYR---VFIGPQNCSCG-----RGTFCHLLFVMLRVFQLEPSPDMLMKRTL 370
QY 419 RFRFANQOVTDALNTSPHMLDLLRLHSSWPQVYGFRLACLCCKVVSASLWGTNRHNRFF 478
DB 371 K-----NFEV-----ESLFQYHSRRSRIRKAPSRN----- 397
QY 479 KNLAKFTSLGKYGLSLQELMWMKMKVEDCHWLRSRSPGKDRVPAAEHLRL---ERILATFL 535
DB 397 -TIQKFS-----RMSNCHTLSS--STSTSSSENSIKDEEQMCPICL 437
QY 536 FWMMDTYVQVLLRFFYITESTFQKNRILFFYRKSQVSKLQSIGVROHLRVLRLSELEE 595
DB 438 LGMLEESLTVCEDCG-----RNKLHHHMSIWA----- 468
QY 596 VRHODTWLAMPICRLR-----IPKNGLRPIVNNMSYSN---GTRALGR 638
DB 469 CHRNRREP-LICPLCRSKVRSHDFVSHELSSPVDSPVTSILRGVQVQSPSPQPVAGSQ---RR 524
QY 639 KQAQHF-----TQRLKTLFSLMNYERTKHPHL--MSSVLGMNDIYRTWPAFVLRVRL 690
DB 525 NCESNFNLTHYGTQIIPPA-----YKDLAEPWIAQAFGMELYGCL-FSRNNVREMAURL 578
QY 691 DQTPRMVFKADVTCA 706
DB 579 SH-----DVSGA 585

```

```

RESULT 6
RTUK_DROFU STANDARD; PRT; 916 AA.
ID RTUK_DROFU
AC P21329;
DT 01-MAY-1991 (REL. 18, CREATED)
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE RNA-DIRECTED DNA POLYMERASE FROM MOBILE ELEMENT JOCKEY (EC 2.7.7.49) (REVERSE TRANSCRIPTASE).
GN POL.
OS DROSOPHILA FUNEBRIS (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1911.1;
RX MEDLINE; 91067678.
RA MIZROKHI L.J., MAZO A.M.;
RT "Evidence for horizontal transmission of the mobile element jockey between distant Drosophila species.";
RT PROC. NATL. ACAD. SCI. U.S.A. 87:9216-9220(1990).

```



RX MEDLINE: 94188926.  
 RA SANDAL N.N., MARCKER K.A.;  
 RT "Similarities between a soybean nodulin, Neurospora crassa sulphate  
 RL permease II and a putative human tumour suppressor.";  
 RL TRENDS BIOCHEM. SCI. 19:19-19(1994).  
 CC -!- FUNCTION: UPTAKE OF SULFATE INTO THE CELL.  
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.  
 CC -!- INDUCTION: HIGHLY EXPRESSED, BUT ONLY IN CELLS SUBJECT TO SULFUR  
 CC -!- LIMITATION, AND IT IS TURNED ON BY THE POSITIVE-ACTING CYS-3  
 CC -!- SULFUR REGULATORY PROTEIN.  
 CC -!- SULFATE PERMEASE II IS MAINLY FOUND IN MYCELIA.  
 CC -!- SIMILARITY: BELONGS TO THE SULFATE PERMEASE FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M59167; G168911; ALT\_SEQ.  
 CC DR PIR: A37956; A37956.  
 CC DR PROSITE: PS01130; Sulfate transp. 1.  
 CC DR PFAM: PF00916; Sulfate transp. 1.  
 CC DR TRANSPORT; TRANSMEMBRANE; GLYCOPROTEIN.  
 CC KW TRANSMEM 71 91 POTENTIAL.  
 CC FT TRANSMEM 103 123 POTENTIAL.  
 CC FT TRANSMEM 128 148 POTENTIAL.  
 CC FT TRANSMEM 171 191 POTENTIAL.  
 CC FT TRANSMEM 193 213 POTENTIAL.  
 CC FT TRANSMEM 271 291 POTENTIAL.  
 CC FT TRANSMEM 326 346 POTENTIAL.  
 CC FT TRANSMEM 363 383 POTENTIAL.  
 CC FT TRANSMEM 451 471 POTENTIAL.  
 CC FT TRANSMEM 474 494 POTENTIAL.  
 CC FT CARBOHYD 23 23 POTENTIAL.  
 CC FT CARBOHYD 578 578 POTENTIAL.  
 CC SEQUENCE 788 AA; 87864 MW; 4BB097A8 CRC32;  
 SQ  
  
 Query Match 1.7%; Score 102.5; DB 1; Length 788;  
 Best Local Similarity 21.6%; Pred. No. 7.2;  
 Matches 93; Conservative 52; Mismatches 170; Indels 115; Gaps 19;  
  
 QY 55 LVCWHGQPPADLSFHQVSSLAKE---LVARYVQLCERNERNVLAFGFELLNEARGG 110  
 DB 254 LVLVQMGKRYPROQRAWFFVSTLRKVFIIILYLVSWLVNKHVDPKKAHKIL---GH 309  
 QY 111 PPMFTSVRSYLPNTVIETLRVSGAMLLLSRVGDDLLVYLLAHCAL-----YLLV 162  
 DB 310 VPSGQHGKAPLNDNEILSA---ISG-----DIPTTILVLLIEHIAISKSFGRVNNYII 360  
 QY 163 PPSQAYQVCG-----SPLXQI-----CATTDIWPVSASYSRTRPVGR 200  
 DB 361 NPSQELVAIGFTNLGPFGLALPRYRIILENHOGQSCSHSRW----HYLRPRPARS 416  
 QY 201 NFTNRLFLQIKSSRQKAPLAPLSRGTRKRLSLTSTSVPSAKKARCYPV----- 253  
 DB 417 LCAHLRLLLHPQOORRRH-DHPRRRRDSHSGVYKFWLTS-----PLEVVIFAG 465  
 QY 253 -----PRVEGPHQVLPPTSGKSW--VPSPAR---SPEVPTAKDLSSKGKVDLSLS 301  
 DB 466 VEVSIFTSIENGIYVTAASGAVLLWRIAKSPGKFLGTOTEIYAPRELVRGSK--DSGRS 523  
 QY 302 GSVCCKHKPSSVSLSPRONAFQLRPFTETRH---FLYSRGDGQERLNPFSLLSNLPQN 358  
 DB 524 RKACRSR-STTAFSLDRDDLSNELQLISTPWFIVYRTEGGLNYVNSAKHLDNLFIH 582  
 QY 359 L---TGARRLVEIIFLSRP-----RTSGPLCRT-----HRLSR 390  
 DB 583 VFKHTRTELNFELKGLDRPNDPGPRPRLPNRRARFAPDPAHHPRLLRQLRRCDDR 642

QY 391 YWQMRPLFOQ 400  
 DB 643 PRLQDLRNQ 652  
  
 RESULT 9  
 T2D5\_HUMAN STANDARD; PRT; 677 AA.  
 ID T2D5\_HUMAN  
 AC P49848;  
 DT 01-OCT-1996 (REL. 34, CREATED)  
 DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
 DE DE TRANSCRIPTION INITIATION FACTOR TFIIID 70 KD SUBUNIT (TAFII-70)  
 DE (TAFII-80) (TAFII80).  
 GN TAF2E OR TAFII70.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;  
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.  
 RN [1]  
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
 RX MEDLINE: 94085406.  
 RA WEINZIERL R.O., RUPPERT S., DYNLACHT B.D., TANESE N., TJIAN R.;  
 RT "Cloning and expression of Drosophila TAFII60 and human TAFII70  
 RT reveal conserved interactions with other subunits of TFIIID.";  
 RL EMBL J. 12:5303-5309(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-PLACENTA;  
 RX MEDLINE: 95396764.  
 RA HISATAKE K., OHTA T., TAKADA R., GUERMAH M., HORIKOSHI M.,  
 RA NAKATANI Y., ROEDER R.G.;  
 RT "Evolutionary conservation of human  
 RT TATA-binding-polyptide-associated factors TAFII31 and TAFII80 and  
 RT interactions of TAFII80 with other TAFs and with general  
 RT transcription factors.";  
 RL PROC. NATL. ACAD. SCI. U.S.A. 92:8195-8199(1995).  
 CC -!- FUNCTION: TFIIID IS A MULTIMERIC PROTEIN COMPLEX THAT PLAYS A  
 CC CENTRAL ROLE IN MEDIATING PROMOTER RESPONSES TO VARIOUS ACTIVATORS  
 CC AND REPRESSORS. BINDS TIGHTLY TO TAFII-250 AND ALSO DIRECTLY  
 CC INTERACTS WITH TAFII-40.  
 CC -!- SUBUNIT: TFIIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A  
 CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs).  
 CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -!- SIMILARITY: BELONGS TO THE TAF2E FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: L25444; G437385;  
 CC EMBL: U31659; G1136306;  
 CC TRANSFAC: T00783;  
 CC TRANSCRIPTION REGULATION: NUCLEAR PROTEIN.  
 CC SEQUENCE 677 AA; 72668 MW; C270C85F CRC32;  
 SQ  
  
 Query Match 1.7%; Score 101; DB 1; Length 677;  
 Best Local Similarity 21.0%; Pred. No. 7.5;  
 Matches 89; Conservative 32; Mismatches 114; Indels 188; Gaps 20;  
  
 QY 7 CPAVRSLS-LESRYREVWPLATFVRRLGPEGRRLVQDPKIVRTLVQAQCLVCMH----- 60  
 DB 311 CIVSRQLCLRPDNDVNWALRDFAR-----LVQA--ICKHESTTN 349  
 QY 60 -----WGSQPPADLSFHQVSSLAKEVAVVQ-----RLCERNERNVLA 101  
 DB 350 NIOSRITKTKFTKSWDEKTPWTRYGSIAGLAELGHVDVTKTLILPRLQOEGER----- 403  
 QY 102 ELLNEARGGPPMAFTSSVRSYLPNTVIETLRVSGAMLLLSRVG--DOLLVYLLAHCA--L 158





501	QY	501	KKMVEDCHWLSSPGKRVPAAEHRLRERLIALFLWMDTYVVO-----LLRSF	550
1440	Db	1440	AKSHVLLATPVQFELLRRWRQRKNIQSLBELMYDDAHETISQGVYGAIVETLSRM	1495
551	QY	551	FYITESTQKNLFFYKSV-----W-----SKLQSIGVROHLERVLRELSSQBEVR	597
1496	Db	1496	IFIATOLEKKIRFVCLSNCLANARDFGEWAGMTKSNYINFSERIEPLEINIOSFDVE	1555
598	QY	598	HHQDTWLAMPICRLRFIPKPNGLRPVNMYSMTGTRALGRKKQAQHTQRLKTL-----	652
1556	Db	1556	HISFNF-----SMLQWAFEAASAAAAGNRSNVFLPSRKDCMEVASA	1597
652	QY	652	-----FSMLNYE-----RUKPHLMGSSVLGM--ND---IYRTWR	681
1598	Db	1598	FMKFSKAIEWMLNVEEQIVPIEKLTDGHLRAPLKGVGILYKGMASNDERIVKRLYE	1657
682	QY	682	AFVLRVRLADQTPRMFYVKADY-----TGAYDAIPQGLVEVWAMIRHSESTYCIROYA	736
1658	Db	1658	YGAVSVLLISKDCSAFACKTDEVIILGTNLYDG-----AEHKYMPYTIINELL	1704
737	QY	737	VYRDSQGVHKSFRQVNTLSLOPYMGQFL-----KHLQ-----DSDSALRNSVVI	785
1705	Db	1705	EMVGLASGNDSWAGKVLIIITSHMKAYKKFLIEPLPTESYLOYYIHTLNEIANSII-	1764
786	QY	786	EQSISMNESSLSLDFLHFLRHVSVKIGDRCYTCQGIPOGSS--LSTLCSLCFGDME	843
1764	Db	1764	-----QSKDCVDWFTSYFYRRIHVNPSTYGVDRTPSHGISVFLSNLV-ETCLNDLV	1815
844	QY	844	NKLFAEVRQDGLLRFVDDFLVTPHL-----DOAKTEFLSTLVHGV--PEYGC-	890
1816	Db	1816	ESSFIELD-----DTEAEVTAEVNGGDDEATEIISTLSNGLASHYGVSYFTTIQSF	1866
890	QY	890	-----MINLQKTVNPPVEPGLTGAA-----PYQLPAH	918
1867	Db	1867	VSSLSSTLTKNMLYVLSTAVEPESVPLRKGDRLIVKLKRLPLRFPEH	1916

RESULT 12

CUL2 HUMAN	STANDARD;	PRT;	745 AA.
ID	CUL2_HUMAN		
AC	Q13617; 000200;		
DT	01-NOV-1997 (REL. 35, CREATED)		
DT	01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)		
DT	15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)		
DE	CULLIN HOMOLOG 2 (CUL-2).		
GN	CUL2.		
OS	HOMO SAPIENS (HUMAN).		
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;		
OC	PRIMATES; CATARRHINI; HOMINIDAE; HOMO.		
EN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=KIDNEY;		
RA	MEDLINE; 97225922.		
RA	PAUSE A., LEE S., WORREL R., CHEN D.Y.T., BURGESS W.H.,		
RT	LINEHAN W.M., KLAUSNER R.D.;		
RT	"The von Hippel-Lindau tumor-suppressor gene product forms a stable		
RT	complex with human CUL-2, a member of the Cdc53 family of proteins.";		
RT	PROC. NATL. ACAD. SCI. U.S.A. 94:2156-2161(1997).		
RN	[2]		
RP	SEQUENCE OF 95-745 FROM N.A.		
RX	MEDLINE; 96279828.		
RA	KIPREOS E.T., LANDER L.E., WING J.P., HE W.W., HEDGECOCK E.M.;		
RT	"Cul-1 is required for cell cycle exit in C. elegans and identifies a		
RT	novel gene family.";		
RL	CELL 85:829-839(1996).		
CC	-!- FUNCTION: FORMS A STABLE COMPLEX WITH THE VHL TUMOR SUPPRESSOR.		
CC	-!- SIMILARITY: BELONGS TO THE CULLIN FAMILY.		

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its contents in no way



Mon Aug 16 10:21:27 1999

entities requires a license agreement (see http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).

CC  
CC  
CC  
DR EMBL; M22874; G157825; -  
DR EMBL; M38643; -; NOT\_ANNOTATED\_CDS.  
DR PIR; JTO396; JTO396.  
DR FLYBASE; FBgn0001283; Jockey.  
DR PFAM; PF00078; rvt; 1.  
KW TRANSFERASE; RNA-DIRECTED RNA POLYMERASE; TRANSPOSABLE ELEMENT.  
SQ SEQUENCE 916 AA; 103430 MW; 6A13C12F CRC32;

Query Match 1.7%; Score 97.5; DB 1; Length 916;  
Best Local Similarity 21.7%; Pred. No. 21;  
Matches 83; Conservative 61; Mismatches 164; Indels 75; Gaps 20;

Qy 665 LMSSVLGMNDIYRTWR-AFVLRVRLADOTPRMYFYKADVTGAYDAIPO-GKLVE--VVA 720  
Db 479 LIFNSVLVDGVGFFKANKASASIIIMHKTGTP---TDVDSYRPTSLPLSLGKIMERLIILN 534  
Qy 721 NMIRHSESTYCIROYAVVRDSOG---QVHKSFRRQVTTLSLDLOPYMGOFKHLQSDAS 777  
Db 535 RLLTCKDVTAKPQFGFRLQHGTPQLHRVNVFALEAMENKEYAVGAFLD-----587  
Qy 778 ALRNSVVEQSI-----SMNESSSLDFFLH-----FLRHSVVKIGRCYTO-----C 821  
Db 587 -----IQAFDRVHPGLLYKAKRLFPPQLYLWVKSFLERTFHVSDGVKSSIKPIA 639  
Qy 822 OGIPQGSSTLLCSLCFGDMENKL-FAEV-ORDGLLRRFVDDFLVTPHLDQAKFLST 879  
Db 640 AGVPGSVGLPTLYSVFASDMPHTPTVTEDEEDVLIATYADD---TAVLTGKSLIAA 695  
Qy 880 LVHGVPEY-----GCMINLOKTV-VNFPVEPTGLGAA--PYQLPAHCLFPWCG 925  
Db 696 -TSGLOEYLDAPQQAENWVNRINAEKANCANTFANRTGSCPGVSLNGRLRHQAAYL 754  
Qy 926 LLLDTQTLVEFCDYSGYATSIKTSFTQSVFKAGKMRNKLVLRLKCHG-----LF 979  
Db 755 IYLDRL--TFSRHITNQAFRTKVARMSWLIA---PRNKLGLCKVNIYKSLIAPCLF 809  
Qy 980 LLOVNSLOTVCINIYKIFLLOA 1002  
Db 810 YGLQVGYAAKS-HLNKIRILOA 831

RESULT 15  
HD\_FUGRU STANDARD; PRT; 3148 AA.  
ID HD\_FUGRU  
AC P51112;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HUNTINGTIN (HUNTINGTON'S DISEASE PROTEIN HOMOLOG) (HD PROTEIN).  
GN HD.  
OS FUGU RUBRIPES (JAPANESE PUFFERFISH) (TAKIFUGU RUBRIPES).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;  
TELEOSTEI; EUTELEOSTEI; ACANTHOPTERYGII; PERCOMORPHA;  
OC TETRAODONTIFORMES; TETRAODONTIDEI; TETRAODONTIDAE; FUGU.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95375788.  
RA BAKENDALE S., ABDULLA S., ELGAR G., BUCK D., BERKS M., MICKLEM G.,  
RA DUBIN R., BATES G., BRENNER S., BECK S., LEHRACH H.;  
RT "Comparative sequence analysis of the human and pufferfish  
Huntington's disease genes";  
CC NAT. GENET. 10:67-76(1995).  
CC -!- FUNCTION: MAY PLAY A ROLE IN MICROTUBULE-MEDIATED TRANSPORT OR  
CC -!- VESICLE FUNCTION.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -!- POLYMORPHISM: THE POLY-GLN REGION (FOUR RESIDUES) DOES NOT APPEAR  
CC -!- TO BE POLYMORPHIC. EXPLAINING THE ABSENCE OF A HD-LIKE DISORDER.  
CC -!- SIMILARITY: CONTAINS 10 HEAT REPEATS.  
CC -!- SIMILARITY: BELONGS TO THE HUNTINGTIN FAMILY.

QY 203 TN-----LRFLOQIK-----SSRQEPKPLPSRGTRKHLSTSTSVPSAKA 247  
Db 885 TAGSAQANLKLQARDWVSTADVSGSLDTR--TIENWGNRRAPPTSGPRS---940  
QY 248 RCYVPRVEEGHRQVLPFSGKSW-----VSPARSPEVPTAEKLSKGVSDLSL 300  
Db 940 -CTP-----GPPGQ--PSPSG--WRPPGGRTPLARRAQ--PPARPATCAGSPQPDVSR 987  
QY 301 SGVCCCKHKPSST-----SLSPPPQNAFQLRPFETETRHFLYSRGDGOERLNPFL 351  
Db 988 AS---CRHAWDARMPVRVGHQSHLSASERRALPERSLHA-HCHYSFPFRAERSGRFL 1043  
QY 352 LSNLOP-----NLTGARRLVEIIFL-----GSRPR-----TSGPL---C 382  
Db 1044 PLFPEPPEPDDLPLGPEQLARREALRAAWARGPRPHASLPSSVAEAFTRSNPLPARC 1103  
QY 383 RTHLSRYWOMRPLFOOLLVNHAEQYVRL 413  
Db 1104 TGHACACPCQSPSCRHV-----AQTOSLRL 1130

RESULT 14  
RTJK\_DROME STANDARD; PRT; 916 AA.  
ID RTJK\_DROME  
AC P21328;  
DT 01-MAY-1991 (REL. 18, CREATED)  
DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)  
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)  
DE RNA-DIRECTED DNA POLYMERASE FROM MOBILE ELEMENT JOCKEY (EC 2.7.7.49)  
DE (REVERSE TRANSCRIPTASE).  
GN POL.  
OS DROSOPHILA MELANOGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;  
OC DROSOPHILIDAE; DROSOPHILA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 89108009.  
RA PRIMAEGLI A.F., MIZROKHI L.J., ILYIN Y.V.;  
RT "The Drosophila mobile element jockey belongs to LINEs and contains  
coding sequences homologous to some retroviral proteins";  
RL GENE 70:253-262(1988).  
RN [2]  
RP SEQUENCE OF 1-184 FROM N.A.  
RX MEDLINE; 87303653.  
RA MIZROKHI L.J., PRIMAEGLI A.F., ILYIN Y.V.;  
RT "Drosophila mobile element jockey is a retroposon and encodes the  
3AG-specific protein sequence characteristic for retroviruses";  
RL DOKL. AKAD. NAUK SSSR 294:1235-1239(1987).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE; 91330871.  
RA IVANOV V.A., MELNIKOV A.A., SIUNOV A.V., FODOR I.I., ILYIN Y.V.;  
RT "Authentic reverse transcriptase is coded by jockey, a mobile  
Drosophila element related to mammalian LINEs";  
KL EMBO J. 10:2489-2495(1991).  
CC -!- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =  
CC N PYROPHOSPHATE + DNA(N).  
CC -!- COFACTOR: MAGNESIUM AND MANGANESE.  
CC -!- ENZYME REGULATION: INACTIVATED BY SULPHYDRYL REAGENT.  
CC -!- PREFERENCES POLY(RC) AND POLY(RA) AS TEMPLATE AND ACTIVATED DNA  
CC IS NOT EFFECTIVE. HAS A TEMPERATURE OPTIMAL OF 26 DEGREES CELSIUS.  
CC -!- SIMILARITY: STRONG, TO THE PUTATIVE REVERSE TRANSCRIPTASE OF  
CC OTHER LONG INTERSPERSED ELEMENTS (LINEs) IN EUKARYOTES.  
CC -!- SIMILARITY: STRONG TO THE EQUIVALENT PROTEIN OF DROSOPHILA  
CC FUNEBRIS.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: X82939; G804981; -  
KW REPEAT.  
FT DOMAIN 148 272 HEAT REPEATS DOMAIN 1.  
FT DOMAIN 701 898 HEAT REPEATS DOMAIN 2.  
FT DOMAIN 1527 1568 HEAT REPEATS DOMAIN 3.  
FT DOMAIN 18 21 POLY-GLN.  
FT DOMAIN 679 682 POLY-ALA.  
FT DOMAIN 1104 1108 POLY-SER.  
SQ SEQUENCE 3148 AA; 348932 MW; CBB3AA6A CRC32;  
  
Query Match 1.6%; Score 97; DB 1; Length 3148;  
Best Local Similarity 19.0%; Pred. No. 1.4e+02;  
Matches 244; Conservative 177; Mismatches 503; Indels 358; Gaps 64;  
  
QY 33 PEGRRVLPQDPKPIYRTLVQAQCLVCMHMGSGPPADLSFHOYSSL--KELVARVQRLCE 90  
DB 1046 PE-RRTLTGTANMVLSSA-----WF-----PDLSSAQDALLCGNLLAAVAPK-CL 1094  
QY 91 RNERNVLAFCFELLNEARGGPPMATSSVRSYLPNTVETLRVSGAWMLLSRVGDDLLV 150  
DB 1095 RN-----PWAGEDSSSSSTNTSGTHKMEBPWAALSADRAVAVVE 1135  
QY 151 YLLAHCALYLLVPPSCAYOV---CGSPLOYCAITDIWPSVASYRTPRGVNFNLR 206  
DB 1136 QLFSHLLKVLNI---CAHVLDTPPGVPVKATLPSTNTPSL-----PIRRKGD----- 1184  
QY 207 FLOQKSSRQAPKP-----LALPSRGTKRHLSTSTSVPS-----AKKARCPVPRVEE 257  
DB 1184 -KDAVDSSAPLSPKNGNEANTGRTESTGSTAVHKSTILGSFYHLPYLYLVKATH 1242  
QY 258 GPHROVLTPPSGKSWPSPARSP-EVPTAEKDLSSGKVSDL--SLSGSV---CCKHKPS- 312  
DB 1243 ANFKVLDLHNSOERGFGLRAALDVLQLELATINDKRCVEILGYLASCFSREPTM 1302  
QY 312 STSLSPRQNAQFURPIETRHFLY--SRGDQGE-RLNPSFLLSNLPNI----- 360  
DB 1303 ATVCVQQLKTLFGTNLASQVEGFLSGPSRSGKALRGS-----SSLRGLYHYCFMAYP 1358  
QY 360 -----TGARRLVEIFLGSRRPTSGPLCRTHRLSRRYW--QMRPLFQQLLVNHAEC 408  
DB 1359 THFTQALADASLRNMVQ-----AEHQDTSGWFDVMOKTSNQLRSNTANA 1403  
QY 409 QYVR-----LLRSHCR-----FRANQOQVTDALNTSPHLMDLRLHSSPWQVYGFRLAC 458  
DB 1404 ARHRGDKNAIHNLIRLFEPLVLIKALKQYTTSTVALORQVLDLLA-----QLVQ-LRVN 1456  
QY 459 LCKVVSASLW-----GTRNRRERFFKNKFFISLKGKYSGLQELMWMK-V 504  
DB 1457 YCLDSQVFIQFVLKOFYEYEVQFRDSEAIIPNIEFFLVLLSYERYHKSQIISIPKII 1516  
QY 505 EDCHLRSSPGK---DRVPAE---HRL-----RERILATFLWMDT 541  
DB 1517 QLCGIGASGRKAVTHAIPALQPIVHDLFVLGRSNKADAGKELETOKEVYVVMLLRLVQY 1576  
QY 542 YVQLLRSFFYITESTQKNRFFYKRSVMSKQSGVROHLERVLRLSQRSEVRHOD 601  
DB 1577 H--QVLEMFILVQOCHKEN-----EDKWRKLS-----ROIADVILPMIAKQOM--HLD 1621  
QY 602 TWLAMPICRLRF-IPKPNGLRPIVNMYSMGTRALGRKQA-----QHFTORLKLTFSM 654  
DB 1622 SPEALGVNLTFTETVAPSSLRPVDMLKSMFTTPVTVMASVATVQLMWVSGILAVLRVLVSQ 1681  
QY 655 LN-----YERTKHPHLMGSSVLGNNDIYRTWRAFLVLRVLRALDQTPRMFYKADVGA 706

DB 1682 STEDIVLSRIHELISLSPHLLSCHT-----IKRLQQPNLSPSDQPDAGDQ 1725  
QY 707 YDAIPQGG-----LVEVANNIRHSESTYCIROYAVVRRDSOGQVHKSPRQ 753  
DB 1726 QNOEPNGEAKSLPETAFARLIQVLGVLLDDISSRH-----VKVDITEQQHTFFCQ 1778  
QY 754 VTTLSDLQPYMGQFLKHLODSASALRNSVIEQISNMNESS--SLFDEF---LHFLR 807  
DB 1779 LGTLL-----MCLIHVFKS--GMFRITVAASRLKLGESGSHGTEFFYPLGLELSMV 1829  
QY 808 H-----SVVKIGD---RCYTQCQIGPOGSSLS--TLLCSLCFGDMENK-- 846  
DB 1830 HCLITTHPSVLLWCQVLLIIDYTNYSWIEVHOTPGHLSLCTKLLSPHSSSGEGEKPE 1889  
QY 846 -----LFAEVORDGILLRFVDDFLLVTPHLDQAKTFTSLVHGVPYEGCMINLQKTVNF 900  
DB 1890 TRLAMINREIVRGALILFC-DYVCQNLHDSHETWL-----IVNHVRDLIDL 1936  
QY 901 PVEPCTLGGAAPYO---LPAHCLFPMCGLLDQTLEVFCDYSGYAQTSIKTSLTFSVF 957  
DB 1937 SHEP-----PVODFISAVHRNSAASGLFI--QAIOQRCD-----NLNSPTML 1976  
QY 958 KAGTMRNKLKLSVLRKCHGLFLDLQVNSLQTVGINY-----KIFLLQAYRFA 1007  
DB 1977 K--KTL--QCLEGIHLSQSGSLLMLYVDKLLSTPFRVLARMVDTLACRRVEMLLAETLQ 2032  
QY 1008 CVIOLPED-----QVRKNLTFFLGIITSSQASCCYAILKYKNPQMTL 1049  
DB 2033 SVALPLEELHRIQOYLOTSGLAQHQRYSLLDRFRATVSDTSS-----PSTPVTSH 2085  
QY 1050 KASGSFPPEAAHWCYQAFLLKLAHSAVYKCLLGP---LRTAQKLLCRKLPEATWTLK 1106  
DB 2086 PLDGDPPPAPELYIADKEWYVALVKSOC---CLHGDVSLLETTE---LLTKLPPADLLSV 2140  
QY 1107 AAD-----PALSTDFQTIL 1121  
DB 2141 SCKEFNLSLLCPLSLGVQRL 2162

Search completed: August 14, 1999, 05:00:13

Job time: 537 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 13, 1999, 15:10:16 ; Search time 18.23 Seconds  
(without alignments)  
3787.820 Million cell updates/sec

Title: US-09-042-460-2  
Perfect score: 5901  
Sequence: 1 MTRAPRCPAVRSLLRSRYRE.....TILKAADPALSTDFQITLD 1122  
Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database : SPTREMBL\_10.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_protist.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5901	100.0	1122	11	070372
2	3505	58.4	1132	4	014746
3	3498	59.3	1132	4	014783
4	1086.5	18.4	523	4	094807
5	624	10.6	989	3	013338
6	621.5	10.5	988	3	013339
7	426.5	7.2	1132	5	076332
8	410.5	7.0	1031	5	000939
9	401.5	6.8	1117	5	077448
10	355	6.0	67	11	035432
11	341	5.8	884	3	006163
12	136.5	2.3	425	2	052231
13	133.5	2.3	561	5	045321
14	131	2.2	778	4	012815
15	129	2.2	502	5	022002
16	125	2.1	1529	10	081068
17	121	2.1	2088	10	023529
18	120.5	2.0	1646	5	019338
19	117	2.0	1569	10	022997
20	116.5	2.0	544	8	099969
21	116	2.0	945	4	094821
22	115.5	2.0	2157	3	000333
23	114.5	1.9	1199	4	014679
24	113.5	1.9	653	11	P97543
25	112	1.9	1963	6	Q28019
26	111.5	1.9	1417	12	Q57631
27	111.5	1.9	2606	12	Q36414
28	110	1.9	907	12	Q69021
29	109.5	1.9	2493	3	Q60055

30	109.5	1.9	710	3	074535	074535 schizosacch
31	108.5	1.8	351	10	Q39492	Q39492 chlamydomon
32	108.5	1.8	1090	11	Q92286	Q92286 mus musculu
33	108.5	1.8	1143	12	Q10896	Q10896 hog cholera
34	107.5	1.8	601	3	Q93956	Q93956 emericella
35	107.5	1.8	3898	12	Q09710	Q09710 pestivirus
36	107	1.8	949	3	Q14093	Q14093 schizosacch
37	107	1.8	1327	4	Q60859	Q60859 homo sapien
38	107	1.8	510	10	Q04210	Q04210 arabidopsis
39	107	1.8	1003	10	Q23218	Q23218 arabidopsis
40	107	1.8	1456	10	Q92794	Q92794 arabidopsis
41	106.5	1.8	542	2	Q68872	Q68872 myxococcus
42	106.5	1.8	1256	10	Q23528	Q23528 arabidopsis
43	106	1.8	584	2	Q08724	Q08724 fremyella d
44	106	1.8	1115	5	Q27764	Q27764 plasmodium
45	106	1.8	2528	5	Q19317	Q19317 caenorhabdi

## ALIGNMENTS

RESULT 1

070372 ID 070372 PRELIMINARY; PRT; 1122 AA.  
AC 070372;  
DT 01-AUG-1998 (Tremblrel. 07, Created)  
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)  
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)  
DE TELOMERASE REVERSE TRANSCRIPTASE.  
GN TERT.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;  
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RA GREENBERG R.A., ALLSOPP R.C., CHIN L., MORIN G.B., DEPINHO R.A.;  
RL Oncogene 0:0-0(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 98393668.  
RA MARTIN-RIVERA L., HERRERA E., ALBAR J.P., BLASCO M.A.;  
RT "Expression of mouse telomerase catalytic subunit in embryos and adult tissues.";  
KW Proc. Natl. Acad. Sci. U.S.A. 95:10471-10476(1998).  
KW EMBL; AF051911; AAC09323.1;  
KW EMBL; AF073311; AAC34821.1;  
KW MGD; MGI:1202709; TERT.  
KW RNA-directed DNA polymerase.  
SQ SEQUENCE 1122 AA; 127977 MW; 222075D6 CRC32;

Query Match 100.0%; Score 5901; DB 11; Length 1122;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MTRAPRCPAVRSLLRSRYREVMPLATEFVRRLGPEGRRLVQGPDKIYRTLVLAQCLVCMHW 50  
Db 1 MTRAPRCPAVRSLLRSRYREVMPLATEFVRRLGPEGRRLVQGPDKIYRTLVLAQCLVCMHW 60  
Qy 61 GSOPPPADLSFHQVSSLSKELVARVQVQLCERNRNVLAFGFELLNEARGPPMAFTSSVR 120  
Db 61 GSOPPPADLSFHQVSSLSKELVARVQVQLCERNRNVLAFGFELLNEARGPPMAFTSSVR 120  
Qy 121 SYLNTVETIETRVSGAWMLLSRVGDDLLVLLAHACALVLLVPPSCAYOVCGSPYQICA 180  
Db 121 SYLNTVETIETRVSGAWMLLSRVGDDLLVLLAHACALVLLVPPSCAYOVCGSPYQICA 180  
Qy 181 TTDIWPVSASVYRTPVGRNFTNLFLOQIKSSSQEAPKPLALPSRGTKRHLSTSTS 240  
Db 181 TTDIWPVSASVYRTPVGRNFTNLFLOQIKSSSQEAPKPLALPSRGTKRHLSTSTS 240  
Qy 241 VPSAKKARCYPPVREVEGPHROVLPTPSGKSWVPSPARSPEVPTAEKDLSSKGKVDLSL 300  
Db 241 VPSAKKARCYPPVREVEGPHROVLPTPSGKSWVPSPARSPEVPTAEKDLSSKGKVDLSL 300



Mon Aug 16 10:21:27 1999

Db	241	VPSAKKACYPVPRVEBPHQVLPFTSGKSWPSPARSPEVPTAEKDLSSKGVSDLSL	300
Qy	301	SGSVCKKPKSTSLSPRONAQOLRPFIEIETRFPLSRGQGRNLPSTLLSNLOPNT	360
Db	301	SGSVCKKPKSTSLSPRONAQOLRPFIEIETRFPLSRGQGRNLPSTLLSNLOPNT	360
Qy	361	GARRIVEIFLGSRRPTSGPLCRHLSRRYQWMPRLFOQLLVNHAECQYVRLRSRHF	420
Db	361	GARRIVEIFLGSRRPTSGPLCRHLSRRYQWMPRLFOQLLVNHAECQYVRLRSRHF	420
Qy	421	RYANQOVTDALNTSPHMLDLRLHSSPWQVYGFRLACLVKVSASLWGTGHRNRRFKN	480
Db	421	RYANQOVTDALNTSPHMLDLRLHSSPWQVYGFRLACLVKVSASLWGTGHRNRRFKN	480
Qy	481	LKFIISLGYKLSLOELMKWKVEDCHWLRSSPGKDRVPAAEHRLRERILATFLWMD	540
Db	481	LKFIISLGYKLSLOELMKWKVEDCHWLRSSPGKDRVPAAEHRLRERILATFLWMD	540
Qy	541	TYVQLLRSFFYITESTFOKNRLFFYKRSWWSKLSIGVROHLRRLRELSOEVRHQ	600
Db	541	TYVQLLRSFFYITESTFOKNRLFFYKRSWWSKLSIGVROHLRRLRELSOEVRHQ	600
Qy	601	DTWLAMPICRLRIPKPNGLRIPVNNYSMTGTRALGRRKOAQHTQRLKTLFSLNIERT	660
Db	601	DTWLAMPICRLRIPKPNGLRIPVNNYSMTGTRALGRRKOAQHTQRLKTLFSLNIERT	660
Qy	661	KHPLMGSSVLGMDIYRTWRAFLVRLALDQTPRMFYKADVTGAYDAIPGKLVVVA	720
Db	661	KHPLMGSSVLGMDIYRTWRAFLVRLALDQTPRMFYKADVTGAYDAIPGKLVVVA	720
Qy	721	NMIRHSESTYCIROYAVVRDSQGVHKSFRROVTTLSLQPYMGQFLKHLQSDASALR	780
Db	721	NMIRHSESTYCIROYAVVRDSQGVHKSFRROVTTLSLQPYMGQFLKHLQSDASALR	780
Qy	781	NSVTEQISMNSSSSLSLDFEFLHRSVVKIGDRCTYOCQIGOGSSLTLLCSLFG	840
Db	781	NSVTEQISMNSSSSLSLDFEFLHRSVVKIGDRCTYOCQIGOGSSLTLLCSLFG	840
Qy	841	DMENKLFPAEVOQDGLLRFVDDFLVTPHLDQAKTFLSTLVHGVPEYGCMLNOKTVNF	900
Db	841	DMENKLFPAEVOQDGLLRFVDDFLVTPHLDQAKTFLSTLVHGVPEYGCMLNOKTVNF	900
Qy	901	PVEPTGLGAAPYQPLPAHCLFPWCGLLDTQTLVEFCDYSYAGTSIKTSITQSVEKAG	960
Db	901	PVEPTGLGAAPYQPLPAHCLFPWCGLLDTQTLVEFCDYSYAGTSIKTSITQSVEKAG	960
Qy	961	KTMRNKLLSVLRKCHGLFLDLQVNSLQTVNCINIKIFLQAYRFHACVQLPQDQVRK	1020
Db	961	KTMRNKLLSVLRKCHGLFLDLQVNSLQTVNCINIKIFLQAYRFHACVQLPQDQVRK	1020
Qy	1021	NLTFEGLIISSOASCYATLKVKNPGMTLKASGSPPEAAHWCYQAFLLKLAHSVIYK	1080
Db	1021	NLTFEGLIISSOASCYATLKVKNPGMTLKASGSPPEAAHWCYQAFLLKLAHSVIYK	1080
Qy	1081	CLLGLPRTAQKLCRLKRLPEATMTILKAAADPALSTDFQITLD 1122	
Db	1081	CLLGLPRTAQKLCRLKRLPEATMTILKAAADPALSTDFQITLD 1122	
RESULT	2		
Q14746			
ID	Q14746	PRELIMINARY;	PRT; 1132 AA.
AC	Q14746;		
DT	01-JAN-1998 (T-EMBLrel. 05, Created)		
DT	01-JAN-1998 (T-EMBLrel. 05, Last sequence update)		
DT	01-NOV-1998 (T-EMBLrel. 08, Last annotation update)		
DE	TELOMERASE REVERSE TRANSCRIPTASE.		
GN	HRT.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
RN	[1]		
RP	SEQUENCE FROM N.A.		

RC	TISSUE=KIDNEY;		
RX	MEDLINE; 97400623.		
RA	NAKAMURA T.M., MORN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,		
RA	LINGNER J., HARLEY C.B., CECI T.R.;		
RT	"Telomerase catalytic subunit homologs from fission yeast and		
RT	human.";		
RL	Science 277:955-959(1997).		
DR	EMBL; AF015950; AAC51672.1;		
KW	RNA-directed DNA polymerase.		
QY	SEQUENCE 1132 AA; 126995 MW; 2DFBEDF3 CRC32;		
QY	Query Match	59.4%;	Score 3505; DB 4; Length 1132;
Db	Best Local Similarity	62.4%;	Pred. No. 8.3e-298;
Db	Matches 719; Conservative 122; Mismatches 260; Indels 52; Gaps 13;		
QY	1	MTAPRCPAYRSLRSRYREVWPLATFVRRILGPEGRRILVQGDPKIYRTLVACQLVCMHW	60
Db	1	MPRAPRCRAVRSLLRSHYREVLPATFVRRILGPEGRRILVQGDPAAFRALVAOQLVCYFW	60
QY	61	GSOPPADLSFHQVSSKELVARVQVORLCERNERNVLAFFGELLNEAGGPPMAFTSSVR	120
Db	61	DARPPAAPSFROVSCLEKELVARVQVORLCERGAKNVLAFFGELLNEAGGPPMAFTSSVR	120
QY	121	SYLPNIVIELRVSGAWMLLSRVGDDLLVLLAHCALYLLVPPSCAYOCVCSPLYQICA	180
Db	121	SYLPNTVTDALRGSGAWMLLSRVGDDLLVLLAHCALYLLVPPSCAYOCVCSPLYQICA	180
QY	181	TTDIPWSVASYSRPTPVGRNFTNLRLQIKSSSQEAPKPLAPSRGTRKHLSTSTS	240
Db	181	ATQARPFPHAS-GPRRLG-----CERAMNHSVREAGVPLGPAPGARRRGGASRS	231
QY	241	VPSAKKACYPVPRVEBPHQVLPFTSGKSWPSPARSPEVPTAEKDLSSK	292
Db	232	LPLKPRRGAAPPEFTPVQGSWNAHPGTRGPDGFCVVSAR-----PAEATSL	286
QY	293	GKVSLSLS-GSVCKKPKSTSLSPRONAQOLRPFIEIETRFPLSRGQGRNLPST	350
Db	287	GALSTRHSHPSVGRQHAGPSTSRPPRWDTPCPVYAEATKHFLYSSGD-REQURPSF	345
QY	351	LLSNLOPNTGARRVELIFLGSRRPTSGPLCRHLSRRYQWMPRLFOQLLVNHAECQY	410
Db	346	LLSSRLPSLTGARRVETIFLGSRRPMPGTPRRLPRLPQRYQWMPRLFELLGNAOCY	405
QY	411	YRLSHCRFTANQOVTDAL-----NTSPHMLDLRLHSSPWQVY	452
Db	406	GYLLKTHCPRAA---VTPAAGVCAKREKPGQSVAAPEEDTPRLVQLLRQHSSPWQVY	462
QY	453	GFLRACLVKVSASLWGTGHRNRRFKNLKFISLGYKLSLOELMKWKVEDCHWLR	512
Db	463	GFVRACLRLVPPGLWGSRRHNRRLNTKFKISLGHAKLSLOELTWKMSVRDCAWLR	522
QY	513	SPGKDRVPAAEHRLRERILATFLWMDTYVQVLLRSFFYITESTFOKNRLFFYKRS	572
Db	523	SPGVGCVPAAEHRLRERILATFLWMDTYVQVLLRSFFYITESTFOKNRLFFYKRS	582
QY	573	KLOSIGVROHLRRLRELSOEVRHQDTWLAMPICRLRIPKPNGLRIPVNNYSMTG	632
Db	583	KLOSIGVROHLRRLRELSOEVRHQDTWLAMPICRLRIPKPNGLRIPVNNYSMTG	642
QY	633	RALGRKQAOHFTORLKTFLFSLNRYERTKPHLMGSSVLGMDIYRTWRAFLVRLALDQ	692
Db	643	RTFRREKRAERLTSRVKALFSLVNIYARRARRPGLGASVLGDDIHRWTFVLRVRAODP	702
QY	693	TPRMFYKADVTGAYDAIPGKLVVVAANRHSESTYCIROYAVVRDSQGVHKSFR	752
Db	703	PPELYFKVDVTGAYDTIPQDRLTEVIASLIK-PQNTYCVRRYAVVQKAAHGVKAFKS	761
QY	753	QVTTLSLQPYMGQFLKHLQSDASALRNSVWIEQISMNSSSSLTLLCSLFLHRLSVYK	812
Db	762	HVSTLTDLPYMQRFVAHLOET--SPLRDVAVTEQSSSLNEASSGLDFVLFPMCHAVR	819
QY	813	IGDRCTYOCQIGOGSSLTLLCSLFLHRLSVYK	872

Db	820	IRGKSYVOCQIPGSGILSTLCSLCYGDNMENKLFAGIRDGLLLRLVDDFLVTPHUTH	879
Qy	873	AKTEFLSTLVHGVPYEGCMINLOKTVNFPVDPOTLGAAPYQUPAHCLFPWCGLLDQT	932
Db	880	AKTEFLTLVRGVPYEGCVNLRKTVNFPVDEALGCTAFVQMPAHGLFPWCGLLDTRT	939
Qy	933	LEVFDYSGYAOITSIKTSLTFQSFVRAKGTMRNKLISVLRKLCCHGLFDLQVNSLQTVCI	992
Db	940	LEVQDSYSSYARTSIRASLTFNRGCFKAGRNRRKLFVGLRKLCHSLFDLQVNSLQTVCT	999
Qy	993	NIYKIFLLOAYRFHACVQLPFPQORVRKNNITFFGLIISSQASCCYAILKVNKPGMTLKAAS	1052
Db	1000	NIYKILLQAYRFHACVQLPFPHQYVKNKPTFFELRVISDTSILCASLYLKAKNAGMSLGAK	1059
Qy	1053	GS--FPPEAAHMLCYQAFELKLAHSAHSVYKCLGLPLRTAOKLRCRKLPEATMTILKAAA	1109
Db	1060	GAAGPLPSEAVQWLCHQAFLLKLTNRVYVPLLGLSLRTAOTQLSRKPLGGTTLTALEAAA	1119
Qy	1110	DPALSTDFQIILD	1122
Db	1120	NPALPSDFKTILD	1132
RESULT	3		
O14783			
ID	O14783	PRELIMINARY;	PRT; 1132 AA.
AC	Q14783;		
DT	01-JAN-1998 (TREMBLrel. 05, Created)		
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)		
DT	01-NOV-1998 (TREMBLrel. 08, Last annotation update)		
DE	TELOMERASE CATALYTIC SUBUNIT.		
GN	HEST2		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
OC	Eutheria; Primates; Catarrhini; Hominidae; Homo.		
RC	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE: 97433088.		
RA	MEYERSON M., COUNTER C.M., EATON E.N., ELLISEN L.W., STEINER P.,		
RA	CADDLE S.D., ZIAUGRA L., BEIJERSBERGEN R.L., DAVIDOFF M.J., LIU Q.,		
RA	BACCHETTI S., HABER D.A., WEINBERG R.A.;		
RT	"HEST2, the putative human telomerase catalytic subunit gene, is		
RL	up-regulated in tumor cells and during immortalization."		
RL	Cell 90:785-795(1997).		
DR	EMBL: AF018167; AAC51724.1;		
SQ	SEQUENCE 1132 AA; 126937 MW; C1E5E2AF CRC32;		

QY	293	GVKSDLSLS-GSVCKCHKPSSSTLSLSPRONAFOALRP-FIETRHFLYSRGDOERLNPSP	350
Db	287	GALSGTRHSHSPVSGRQHAGPSTSRPPRWDTCCPPVYAEKTHFLVSSGD-KEQLRPSF	345
QY	351	LLSNLOPNTGARRLVEIFIGFSRPTSGPLCRTHRLSRRYQWRPLFOQLLVNHAECOY	410
Db	346	LSSSRPSLTGARRLVEIFIGFSRPMWPGTPRRRLPRLQRYQWRPLFLELLGNHAQCY	405
QY	411	VRLLSHCRFRFANOQVTDAL-----NTSPHLDLRLHSSPWQY	452
Db	406	GVLKTHCPLR---VTPAGVCAREKPGQSVAAPEEDTDPRRLVLQLRQHSWPQY	462
QY	453	GFLRACLKVVYSASLWGTGRHNERREFFKNLKFFISLGRYKLSLQELWKMVKVDDCHWLS	512
Db	463	GFVRACLARLVPPGLWGRSRHNERREFLNKKFISLGRKAKLSLQELTWKMSYRGCAWLR	522
QY	513	SPGKDRVPAAEHRLRERLTATLEFLWMDTYVOLLRSFFYITESTQKNRLFFYRKSVWS	572
Db	523	SPGVCVPAAEHRLREELAKFLHLMVSVYVELLSRFFYVTTFTQKNRLFFYRKSVWS	582
QY	573	KLOSTGVQHLERVLRELGOEEVYRHHQDTWLAMPICRLRFIPKPNGLRPIVNMYSMTG	632
Db	583	KLOSIGIRQHLKRVQLRELSEAEVRQREARPALLTSRLREFIPKPDGLRPIVMDYVGA	642
QY	633	RALGRKQAOHETORLKTLSMLNERTKPHLMGSSVIGMNDIYTRAFVLRVRLDQ	692
Db	643	RTFREKRAERLTSRVKALFSVLNERARRPGLGASVIGLDDIIRAWFTFLVLRVRAQDP	702
QY	693	TPRMFYKADVTGAYDAIPOGKLVVYVANNIRHSESTYCIROYAVVRRDSOGVHKFSRR	752
Db	703	PPLEYFKVDVTGAYDTIPQDRLTEVIASTIK-PQNTYCVRYAVVQAAGHGVKFAFS	761
QY	753	QVTTLSLDQPMGOFKHLQSDASALRNSVIEQSIWNSSSSSLDFDFFLHLRHSYVK	812
Db	762	HVSTLTDQPMRQFVAHQLOET--SPLRDAVIEQSSSLNEASSGLFDVFLRFMCHHAVR	819
QY	813	IGDRCYTCOGIGPOGSSLSLTLLCSLCFGDMENKFLAEVORDGLLLRFDVDFLVTPHLDQ	872
Db	820	IRGKSYVQCQIGIPOGSTLTLCSLCYCDMENKFLAIFRQGLLRLVDDFLVTPHLTH	879
QY	873	AKTFLSLTVHGVPYEGCMINLQTVYVPEPCTLGGAPYOLPAHCLFPMCGLLDQTQ	932
Db	880	AKTFLRLTVHGVPYEGCVNLRITVYVPEDEALGGTAFVQMPAHGLFPMCGLLDTRT	939
QY	933	LEVFCDSYGAOTSIRKSTLTFQSVFKAGTKMNKLLSVLRKLCGLFLDLOVNSLQVCI	992
Db	940	LEVQSDSYSVARTSIRASLTFRNGFKAGNMRKLFGLVLRKCHSLFLDLOVNSLQVCT	999
QY	993	NIYKIFLQAYRHACVIOLPDQVRKNKLTFFLGIISQASCCYAILKVKNPQMTLKAS	1052
Db	1000	NIYKILLQAYRHACVIOLPFHQQVWKNPTFFLRVISTASLCYSILKAKNAGMSLGAK	1059
QY	1053	GS---FPPEAAHWLCYQAFLLKLAHVSIVYKCLLGLPLRTAQKLCRLPEATWTLKAAA	1109
Db	1060	GAAGPLPSEAVQWLCHQAFLLKLTTHRVTYVPLGSLRTAQKLRKPGTTLTALEAAA	1119
QY	1110	DPALSTDFQTLTD 1122	
Db	1120	NNALPSDFKTILD 1132	
RESULT	4		
Q94807			
ID	Q94807	PRELIMINARY:	PRT; 523 AA.
AC	Q94807;		
DT	01-MAY-1999	(TremBLrel. 10, Created)	
DD	01-MAY-1999	(TremBLrel. 10, Last sequence update)	
DT	01-MAY-1999	(TremBLrel. 10, Last annotation update)	
DE	TELOMERASE	TRANSCRIPTASE (FRAGMENT).	
GN	HTERT		
OS	Homo sapiens	(Human).	
OC	Eukaryota;	Metazoa;	Chordata;
OC	Eutheria;	Primates;	Catarrhini;
OC			Hominidae;
OC			Mammalia;
OC			Homio.

Mon Aug 16 10:21:27 1999

```

RN SEQUENCE FROM N.A.
RP TAKAKURA M., KYO S., KANAYA T., HIRANO H., TAKEDA J., YUTSUDO M.,
RA INOUE M.; and characterization of human telomerase catalytic subunit
RT "Cloning and characterization of human telomerase catalytic subunit
RT (hTERT) gene promoter."
RL Cancer Res. 0:0-0(1999).
DR EMBL: AB016767; CAB19461.1; -.
KW RNA-directed DNA polymerase.
FT NON_TER 523 523
SQ SEQUENCE 523 AA; 56555 MW; 0F88F1C8 CRC32;

Query Match 18.4%; Score 1086.5; DB 4; Length 523;
Best Local Similarity 48.7%; Pred. No. 1.1e-86;
Matches 269; Conservative 45; Mismatches 171; Indels 67; Gaps 16;

QY 1 MTRAPCPAVSLRLSRVREVWPLATVRLGPEGRRLVQGPDKIYRTILVAOCLVCMHW 60
DB 1 MPRAPCRVAVSLRLSRVREVWPLATVRLGPEGRRLVQGPDKIYRTILVAOCLVCMHW 60
QY 61 GSQPPPADLSFHOYSSSLKELVARVORLCEARNERNVLAFFELLNEARGGPPMAFTSSVR 120
DB 61 DARPPAPSPFQVYSSCLKELVARVORLCEARNERNVLAFFELLNEARGGPPMAFTSSVR 120
QY 121 SYLNTVITLVRSGAWMLLSRVGDDLLVYLAHCALYLLVPPSCAYQVCGSPLYOICA 180
DB 121 SYLNTVITLVRSGAWMLLSRVGDDLLVYLAHCALYLLVPPSCAYQVCGSPLYOICA 180
QY 121 SYLNTVITLVRSGAWMLLSRVGDDLLVYLAHCALYLLVPPSCAYQVCGSPLYOICA 180
DB 121 SYLNTVITLVRSGAWMLLSRVGDDLLVYLAHCALYLLVPPSCAYQVCGSPLYOICA 180
QY 181 TTDIWPVSASVTRPVGRNFTNLRLQOIKSSSQEAPKPLALPSRGTKRHLSTLS 240
DB 181 TTDIWPVSASVTRPVGRNFTNLRLQOIKSSSQEAPKPLALPSRGTKRHLSTLS 240
QY 181 ATQARPPPHAS-GPERRLG-----CERAWNHSVREAGVPLGLPAPGARRRGSASRS 231
DB 181 ATQARPPPHAS-GPERRLG-----CERAWNHSVREAGVPLGLPAPGARRRGSASRS 231
QY 241 VPSAKKARCPVPRVEEGP-----HRQVLPSPGSKW-VPSPARSPVPTAEKDLSSK 292
DB 241 VPSAKKARCPVPRVEEGP-----HRQVLPSPGSKW-VPSPARSPVPTAEKDLSSK 292
QY 232 LPLKPRPRGAPEPPTPVGGSWAHGPGTRGSDRGFCVVSAP-----PAEATSL 286
DB 232 LPLKPRPRGAPEPPTPVGGSWAHGPGTRGSDRGFCVVSAP-----PAEATSL 286
QY 293 GRVSDLSLS-GVCCCKHPSTSLSPRQNAFQLRP-FIETRIFLYSGDQERLNPSPF 350
DB 293 GRVSDLSLS-GVCCCKHPSTSLSPRQNAFQLRP-FIETRIFLYSGDQERLNPSPF 350
QY 287 GALSSTRHSPVSGRQHAGPSTSRPDPWDTPCPVPVYAEKHFYSSGD-KEQLRPSF 345
DB 287 GALSSTRHSPVSGRQHAGPSTSRPDPWDTPCPVPVYAEKHFYSSGD-KEQLRPSF 345
QY 351 LLSNLQPNLTGARRLVEIFIGSRPTSG-----PLC--RHRLSRRYWQMRPLFQQLLV 403
DB 351 LLSNLQPNLTGARRLVEIFIGSRPTSG-----PLC--RHRLSRRYWQMRPLFQQLLV 403
QY 346 LLSLRPLITGARRLVEIFIGSRPTSG-----PLC--RHRLSRRYWQMRPLFQQLLV 405
DB 346 LLSLRPLITGARRLVEIFIGSRPTSG-----PLC--RHRLSRRYWQMRPLFQQLLV 405
QY 404 NHAEOYVLLSHCRFTANOQVTDALNTSPHMLDLRLHSS---PWQVYGLRACLC 460
DB 404 NHAEOYVLLSHCRFTANOQVTDALNTSPHMLDLRLHSS---PWQVYGLRACLC 460
QY 406 G--CSSRTAR--CELRSPQPVSV-VPGRSPALWRPRTQTQPVAV-----CSC 451
DB 406 G--CSSRTAR--CELRSPQPVSV-VPGRSPALWRPRTQTQPVAV-----CSC 451
QY 461 KVPASLWGT-----HNERFFKNLKKFISLGKYSGLQSLQELMKM 502
DB 461 KVPASLWGT-----HNERFFKNLKKFISLGKYSGLQSLQELMKM 502
QY 452 SASTAPLAGVRLRAGLPAPAGAPRPLGLQAQRRLRLNKKFISLGKYSGLQSLQELMKM 511
DB 452 SASTAPLAGVRLRAGLPAPAGAPRPLGLQAQRRLRLNKKFISLGKYSGLQSLQELMKM 511
QY 503 KVEDCHWLRRSP 514
DB 503 KVEDCHWLRRSP 514
QY 512 SVRDCAWLRRSP 523
DB 512 SVRDCAWLRRSP 523

RESULT 5
OL3338 PRELIMINARY; PRT; 989 AA.
AC OL3338;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE TELOMERASE REVERSE TRANSCRIPTASE 1.
GN TET1.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-972H-;

```

```

RX MEDLINE; 97400623.
RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,
RA LINGNER J., HARLEY C.B., CECCH T.R.;
RT "Telomerase catalytic subunit homologs from fission yeast and
RT human."
RL Science 277:955-959(1997).
DR EMBL: AF015783; AAC49803.1; -.
DR PFAM: PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 989 AA; 116456 MW; FDE74202 CRC32;

Query Match 10.6%; Score 624; DB 3; Length 989;
Best Local Similarity 23.7%; Pred. No. 8.3e-46;
Matches 255; Conservative 194; Mismatches 381; Indels 244; Gaps 46;

QY 76 SKELVARVQVQLCERN---ERNVLAFFGELLNE-ARGGPPMAFTSSVRSYLPVITFL 131
DB 76 SKELVARVQVQLCERN---ERNVLAFFGELLNE-ARGGPPMAFTSSVRSYLPVITFL 131
QY 83 SQSELIANYVKOMFDESFEFRRNLLMKGFSMNHEDFRAMHVGQNDLVSTFPVYLISIL 142
DB 83 SQSELIANYVKOMFDESFEFRRNLLMKGFSMNHEDFRAMHVGQNDLVSTFPVYLISIL 142
QY 132 RVSGAWMLLSRVGDDLLVYLAHCALYLLVPPSCAYQVCGSPLYOICATTDINPVSAS 191
DB 132 RVSGAWMLLSRVGDDLLVYLAHCALYLLVPPSCAYQVCGSPLYOICATTDINPVSAS 191
QY 143 E-SKNWOLLLEIIGSDAMHYLLSKGSEIPEALPNDNYLQISGIPLEK----- 188
DB 143 E-SKNWOLLLEIIGSDAMHYLLSKGSEIPEALPNDNYLQISGIPLEK----- 188
QY 192 YRTPRVGRNFTNLRLQOIKSSSQEAPKPLALPSRGTKRHLSTLSV-----PS 243
DB 192 YRTPRVGRNFTNLRLQOIKSSSQEAPKPLALPSRGTKRHLSTLSV-----PS 243
QY 188 -----NNV--FEETVSKRRTIETSIQNKARKEVSWNSISISFISFYRSS 234
DB 188 -----NNV--FEETVSKRRTIETSIQNKARKEVSWNSISISFISFYRSS 234
QY 244 AKKARCPVPRVEEGPHROVLPSTGKSNVPSPARSPVPTAEKDLSSKGKVDLSLS-G 302
DB 244 AKKARCPVPRVEEGPHROVLPSTGKSNVPSPARSPVPTAEKDLSSKGKVDLSLS-G 302
QY 235 YKAFK-----QDLYFNHL 247
DB 235 YKAFK-----QDLYFNHL 247
QY 303 SVCCCKHPSTSLSPRO---NAPQLPPTIETRHFLYSGDQERLNPSELLNLQPN 358
DB 303 SVCCCKHPSTSLSPRO---NAPQLPPTIETRHFLYSGDQERLNPSELLNLQPN 358
QY 248 SICDRNTVMQWIFPQFGLINAFQVQL-----HKVPLVVSQSTVFP- 293
DB 248 SICDRNTVMQWIFPQFGLINAFQVQL-----HKVPLVVSQSTVFP- 293
QY 359 LTGARLVEIFIGSRPTSGPLCRTHLSRRYWMQMRPLFQQLLVNHAEOYVRLRSHC 418
DB 359 LTGARLVEIFIGSRPTSGPLCRTHLSRRYWMQMRPLFQQLLVNHAEOYVRLRSHC 418
QY 293 -----KRLKVV-----PLIEQTAKRHLRIS-----LSKVNH-YCPVI- 326
DB 293 -----KRLKVV-----PLIEQTAKRHLRIS-----LSKVNH-YCPVI- 326
QY 419 RFRTAQOQVTDALNTSPHMLDLRLHSSFPQVQVGLRACLCVKVVSASLWGTTHNRRF- 478
DB 419 RFRTAQOQVTDALNTSPHMLDLRLHSSFPQVQVGLRACLCVKVVSASLWGTTHNRRF- 478
QY 326 --DTHDDE-----KILSYSLAPNQVFAELRSILRVFPKPLWG---NORIFE 367
DB 326 --DTHDDE-----KILSYSLAPNQVFAELRSILRVFPKPLWG---NORIFE 367
QY 478 --FNALKKPFISLQYKLSLOELMKMKVDECHWL-----RSSPGKDRVPAARHLRERIL 531
DB 478 --FNALKKPFISLQYKLSLOELMKMKVDECHWL-----RSSPGKDRVPAARHLRERIL 531
QY 368 IILKDLTFLLKLSRYESFSLHYLMSNLIKISEIWLVLGRSNA---KMCLSDFEKRKQIF 424
DB 368 IILKDLTFLLKLSRYESFSLHYLMSNLIKISEIWLVLGRSNA---KMCLSDFEKRKQIF 424
QY 532 ATFLFLMDTYVQVQLLSRFFIITESTFQKNRFLFYKRVKWSKLOSGVROHLEVRVL--- 589
DB 532 ATFLFLMDTYVQVQLLSRFFIITESTFQKNRFLFYKRVKWSKLOSGVROHLEVRVL--- 589
QY 425 AEFYLYNSFIPILOQSFYITESSDLNRKTVFRRDKWKL---CRPFIITSMKEAF 480
DB 425 AEFYLYNSFIPILOQSFYITESSDLNRKTVFRRDKWKL---CRPFIITSMKEAF 480
QY 589 RELSQEVRHHQDTW-LAMPICRLRFPKPNGLRPVNMYSYSGMTRALGRKQAQHT-Q 646
DB 589 RELSQEVRHHQDTW-LAMPICRLRFPKPNGLRPVNMYSYSGMTRALGRKQAQHT-Q 646
QY 481 EKINENNVR--MDTKTTLPPAVIRLLPKKNTFRLLNLRKRLIKOMGSKMLVSTNQ 538
DB 481 EKINENNVR--MDTKTTLPPAVIRLLPKKNTFRLLNLRKRLIKOMGSKMLVSTNQ 538
QY 647 RLKTLFSLMNVETKHPHLMG--SSVLGMN-DIYR---TWRAFVLRVRLDQTPRMVFK 700
DB 647 RLKTLFSLMNVETKHPHLMG--SSVLGMN-DIYR---TWRAFVLRVRLDQTPRMVFK 700
QY 539 TLRPVASILK-----HLNESSGIPFNLEVYMKLLTFKDLKLRHMEGR--KRYFVR 589
DB 539 TLRPVASILK-----HLNESSGIPFNLEVYMKLLTFKDLKLRHMEGR--KRYFVR 589
QY 701 ADVTGAYDAIPQGLVEVVMNIRHSESTCYROYAVVRDSSQGVHKSFRQVTTLSDL 760
DB 701 ADVTGAYDAIPQGLVEVVMNIRHSESTCYROYAVVRDSSQGVHKSFRQVTTLSDL 760
QY 590 IDIKSCYDRIKODLMFRIVKKKKLDPE--FVIRKYATIHATSD-RATKNFVSEAFSFD 646
DB 590 IDIKSCYDRIKODLMFRIVKKKKLDPE--FVIRKYATIHATSD-RATKNFVSEAFSFD 646
QY 761 QPY--MGQFLAKHLOSDASALRNSVWTEQISIMNESSSLDFDLFLHFLRHSVVKIGR 818
DB 761 QPY--MGQFLAKHLOSDASALRNSVWTEQISIMNESSSLDFDLFLHFLRHSVVKIGR 818
QY 647 VPEFVKVQLLS-MKTS-----TLFVDFVDYWTKSSSEIFKMLKEHLSGHIVKIGNSQY 699
DB 647 VPEFVKVQLLS-MKTS-----TLFVDFVDYWTKSSSEIFKMLKEHLSGHIVKIGNSQY 699
QY 819 TQCOGIPQGSLSLTLLCSLCFGDMENKLFQVQDQD-LLLRFVDDFLVTPHLOQAKTEL 877
DB 819 TQCOGIPQGSLSLTLLCSLCFGDMENKLFQVQDQD-LLLRFVDDFLVTPHLOQAKTEL 877
QY 700 LQKVGIPQGSLSLSPFLCHFYMEDLIDELYSFTKKKGSVLLRVDVDFLFTVYNNKDKAKFL 759
DB 700 LQKVGIPQGSLSLSPFLCHFYMEDLIDELYSFTKKKGSVLLRVDVDFLFTVYNNKDKAKFL 759
QY 878 STLHVGVPEYGCMLNQKTVVNFVPEPGTIGGAAPYQLPAHCLFPWCGLLDDTQTLVEF- 937
DB 878 STLHVGVPEYGCMLNQKTVVNFVPEPGTIGGAAPYQLPAHCLFPWCGLLDDTQTLVEF- 937

```

Db 760 NLSLRGFEKHNFTSLEKTVINFNENGINNT--FENESKRMPPFGFVSNRSLDTLL 817  
QY 937 ----CDYSGYAQTSIK-TSLTFQSVFAGKTRMKNKLLSVLRKCHGLFLDQVNSLQTV 991  
Db 818 ACPRIDEALFNSTSVELTKHMGKFFY--KILRSSLASFAQ-----VFIDITHNSKFN 870  
QY 992 INIYKI-----FLQAYRFHACVQLP-----FDQVRKNLTFFLGISSQASC 1035  
Db 871 CNIRYLGYSMCMRAQAYLKRKMDIFIPQRMFIDLLNVIGRKIKWKLAEILGITSRFLS 930  
QY 1036 CYALKVKNPMT--LKASGSFPEAAHMLCYQAFLLKLAHSHVVIYKLLGPLR 1087  
Db 931 SAEVKNLFCGLMRDGLAPSKYHP-----CFEQLIYQFQSLTDLIK-----PLR 974  
RESULT 6  
O13339 ID O13339 PRELIMINARY; PRT; 988 AA.  
AC O13339;  
DT 01-JAN-1999 (TrEMBLrel. 09, Created)  
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)  
DE TELOMERASE REVERSE TRANSCRIPTASE 1 (EC 2.7.7.-)  
GN TRT1 OR SPC29A3.14C.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Archiascomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RX MEDLINE; 97400623.  
RA NAKAMURA T.M., MORIN G.B., CHAPMAN K.B., WEINRICH S.L., ANDREWS W.H.,  
LINGNER J., HARLEY C.B., CECCH T.R.;  
RT "Telomerase catalytic subunit homologs from fission yeast and  
human";  
RL Science 277:955-959 (1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972;  
RA LYNE M., RAJANDREAM M.A., BARRELL B.G., VOLCKAERT G.;  
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THIS IS THE CATALYTIC SUBUNIT AND IS DIRECTLY INVOLVED  
CC -!- SUBCELLULAR LOCATION: NUCLEAR.  
DR EMBL; AF015783; AAC49802.1; -;  
DR EMBL; AL022299; CAA18391.1; -;  
DR PFAM; PF00078; rvt; 1.  
KW Transferase; RNA-directed DNA polymerase; Telomere; Nuclear protein;  
SQ DNA-binding.  
QY SEQUENCE 988 AA; 116328 MW; 39C385A7 CRC32;  
Query Match 10.5%; Score 621.5; DB 3; Length 988;  
Best Local Similarity 23.6%; Pred. No. 1.4e-45;  
Matches 253; Conservative 193; Mismatches 384; Indels 243; Gaps 45;  
QY 76 SKELVARVVQRLCERN-----ERNVLAFGFELLNE-ARGPPMAFTSSVRSYLPNTVIETL 131  
Db 83 SQSELIANVVKQFDSFERRNLLKMGFSMNHEDFRAMHVNGVQNDLVSTFPNVLISIL 142  
QY 132 RVSGAWMLLSRVGDDLLVYLHACALYLLVPPSCAYQVCGSPLYOICATTDIWPVSAS 191  
Db 143 E-SKNWQLLEITIGSDAMHYLLSKGSIFEALPDNLYLOISIPLFK----- 188  
QY 192 YRPTRPVGRNFTNLFLOQIKSSSRQEPAPLPSRCTKRLSLTSTV-----PS 243  
Db 188 -----NNV--FEETVSKKKRTIETSIQNKSAKEVSNWSISISRFSIFRSS 234  
QY 244 AKKARCYVPVRVEGPHROVLPTPSGKSWPSPARSPEVPTAEKDLSSKGVSDLSLS-G 302  
Db 235 YKKEK-----ODLYFNHLH 247

QY 303 SVCCCKHPSPSTLSLSPRQ-----NAFQLRPFIEIHRFLYSGDQERLNPSPFLSLNLPN 358  
Db 248 SICDNTVHMVLOWIFPRQGLINAFQVKQL-----HKVPLVYSQSTVVP- 293  
QY 359 LTGARLVEIIFLGSRRPTSGPLCRTHRLSRRYWQMRPLFQQLLVNHAECOVRLRSHC 418  
Db 293 ----KRLKVV-----PLIQOTAKRLHRIS-----LSKVYNH-YCPYI----- 326  
QY 419 RERTANQOVDALNTSPPHLMDLLRLHSSPWQYGFELRACLCKVVSASLNGRHRERF- 478  
Db 326 --DTHDDE-----KILSYSLKPNQVFAELRSILVRFPKILWG---NORIFE 367  
QY 478 --FKNLKKFTISLKGKLSLOELMMKMKVEDCHWL-----RSSPGKDRVPAAPHLRERIL 531  
Db 368 IILKDLTFLKLSRYESFSLHYLMSNIKISEIWEVLVGLKRSNA---KMLCSDFEKKRQIF 424  
QY 532 ATFLFWLMDTVVQLLRSPFFYITESTFOKNRLLFFYKRSVWSKQSIGVROHLERVRL--- 589  
Db 425 AEFYLYNSFIPILOSFYITESSDLRNTVYERKDIWLL-----CRPFIITSMKMEAF 480  
QY 589 RELSQEVRHHQDTW-LAMPICRLRFIPKPNGLRPIVNMYSMTGRALGRKQAOHFTQR 647  
Db 481 EKINENNVY--MDTQKTLPPAVIRLLPKKNTFRLLNLRKRLIKMGSNKKMLVSTNQT 538  
QY 648 LKTLFSLMYERTKHPHLMG--SSVLGMN-DIYR---TWRAFVLRVLRALDOTPRYFVKA 701  
Db 539 LRPVASILK-----HLINESGIPLENLEVYMKLLTFKDKLLKRMFGR--KKYFVRI 589  
QY 702 DVTGAYDAIPQGLVEVYVANNIRHSESTYCIQYAVYVRDSQGVHKSFRROVTTLSLQ 761  
Db 590 DIKSCYDIKODLMFRIVKKKLDPE--FVIRYATIHATSD-RATKNFVSEAFSYEDMW 646  
QY 762 PY--MGQELKHLQSDASALRNSVYIQSISMNNESSSLFDFELHLRLSHVVKIGRCYT 819  
Db 647 PFEKVQVLLS-MKTS-----TLFVDFVDYWTYKSSSEIFPKMKEHLSGHVKNIGNSOYL 699  
QY 820 QCOGIPOGSSSLTLLCSLFCGDMENKLFABEVQDQ--LLLRVDFDLVLTPLHDQAKTFLS 878  
Db 700 QKVGIPOGSSILSFLCHFYMEDLIDEYLSFTKKKGVLLRVVDDFLFIVNKKDAKFLN 759  
QY 879 TLVHGVEYGCINLQKTVVNFVEPGTILGGAAPYOLPAHCLFPPWCGLLDQTOTLEVF-- 937  
Db 760 LSLRGFEKHNFTSLEKTVINFNENGINNT--FENESKRMPPFGFVSNRSLDTLLA 817  
QY 937 ---CDYSGYAQTSIK-TSLTFQSVFAGKTRMKNKLLSVLRKCHGLFLDQVNSLQTVCI 992  
Db 818 CPKIDALFNSTSVELTKHMGKFFY--KILRSSLASFAQ-----VFIDITHNSKFN 870  
QY 993 NIYKI-----FLQAYRFHACVQLP-----FDQVRKNLTFFLGISSQASC 1036  
Db 871 NIYRLGYSMCMRAQAYLKRKMDIFIPQRMFIDLLNVIGRKIKWKLAEILGITSRFLS 930  
QY 1037 YALKVKNPMT--LKASGSFPEAAHMLCYQAFLLKLAHSHVVIYKLLGPLR 1087  
Db 931 AEVKNLFCGLMRDGLAPSKYHP-----CFEQLIYQFQSLTDLIK-----PLR 973  
RESULT 7  
O76332 ID O76332 PRELIMINARY; PRT; 1132 AA.  
AC O76332;  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DE TELOMERASE REVERSE TRANSCRIPTASE.  
GN TERT.  
OS Oxytricha trifallax.  
OC Eukaryota; Alveolata; Ciliophora; hypotrichs; Stichotrichida;  
OC Oxytrichidae; Oxytricha.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 98337940.



```

QY 1008 CVIQLPFDORVRKNLTFGLTISSQASCCYAILK 1041
Db 931 CAKE--YKDFKKNLMSMIDLEYSKIIVSTR 962

RESULT 9
O77448
ID O77448 PRELIMINARY; PRT; 1117 AA.
AC O77448
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DE TELOMERASE REVERSE TRANSCRIPTASE.
GN TERT.
OS Tetrahymena thermophila.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B7;
RX MEDLINE; 98337940.
RA BRYAN T.M., SPERGER J.M., CHAPMAN K.B., CECI T.R.;
RT "Telomerase reverse transcriptase genes identified in Tetrahymena
thermophila and Oxytricha trifallax."
PL Proc. Natl. Acad. Sci. U.S.A. 95:8479-8484(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98337941.
RA COLLINS K., GANDHI L.;
RT "The reverse transcriptase component of the Tetrahymena telomerase
ribonucleoprotein complex."
PL Proc. Natl. Acad. Sci. U.S.A. 95:8485-8490(1998).
DR EMBL; AF062652; AAC39135.1;
DR EMBL; AF061284; AAC39140.1;
KW RNA-directed DNA polymerase.
SQ SEQUENCE 1117 AA; 133317 MW; 1B3DF5A7 CRC32;

Query Match 6.8%; Score 401.5; DB 5; Length 1117;
Best Local Similarity 21.2%; Pred. No. 2.9e-26;
Matches 173; Conservative 147; Mismatches 358; Indels 139; Gaps 28;

QY 296 SLSLSGSCVCKHKPSTSLSPFPQNAFQLPFTTHFYLSRGDGE----- 345
Db 190 NFNMGKATSNNONNANLSEKKEQNOYIYPEIQRSQIFCYCNHMGREGVFKSSFFNY 249
QY 345 -RLNPSFLSNLQNLGARRLVEIIFLGSRPRT--SGPLCRTHRLSPRY--WQMRPLF- 399
Db 250 SEIKKGQFQVIOEKLOGRQFINSKIRPDHPQIIKTLKEQY-SKNFSQCEERDLFL 308
QY 399 ---QOLLVNHAEQVRLRLRSHCF-----RTANQOQVTDALNTSPPH 437
Db 309 ETEKIVQNFNINFNINLLKFKCKLPENYQSLKSVQKQVQSENKANOQSCENLFS--- 366
QY 438 LMDLLRLHSSPWQYGFRLACLCCKVWSASLWGTNRNRRERFKNLKFIISLGKYGKLSLQE 497
Db 366 ---LYDTEISYKQITNFRLLQIIONCVNQLLG-KNKFVLEKLYEYVQMKFENOKVLD 421
QY 498 LMKMKVYEDCHWLKSSPGKRVPAEH-RLRERILATFLFLMDTVVQVLLRSFFVITES 556
Db 422 YICFMDVDFVEMFVLDKNQKRTQKKYISDKRTILGLIVFIINKIVIPVLRNFYITEK 481
QY 557 TFQKNRLLFFYKRSVWSKLSQISGVHRLERLRLSELSEEVRRH---ODTWLAMPICRLRF 613
Db 482 HKESQIFYYRKPWKVLSKTI-----VKLEENLEKVEERKLIPEDSFQKYPQCKLRI 535
QY 614 IPRNGRLRPVNNYSNGTRALGRKQQAQHTQRLKTLF--SMLNVERKH--PHLMGSS 669
Db 536 IPKKGSRPMTFL-----RKDKQKNKLNQLMDSQVLFNKLMDLQKIGYS 586
QY 670 VLGNNDIYRTWRAFLVRLVALDQTPWYFVKADVTGAYDAIPGKLIVEVYANNIRHS--- 727

```

```

Db 587 VFDNKOISEKFAQFIEKWKNKGR-PQLYYVTLDIKKCYDSIDQKLL-----NFFNQSDLI 641
QY 727 ESTYCIROYAVVRDSDSQGVHKSFRQVTTLSDLQ-----PYM-----G 765
Db 642 QDTYFINKYLLFQRNKRPLLQIQOQTNNSAMEIEEEKINKPKFKMDINFPYFNLKER 701
QY 766 QFLKHLQSDASAL-----RNSWIEQSISSMNESSSLDFLHFLHRSVYKIG 814
Db 702 QIAYSLYDDDDQILOKGFKEIQSDDRPFIVINQDKPCITKDIHNLKHSIVNVSFN 761
QY 815 DRCYTQCOGIPQSSLSLTLSCFGDMENK-----LFAEYQVQDQ---LLRFRVDDFLVLT 867
Db 762 KVKFRQKRGIPQGNISGLCVSYFGKLEEBYQFLKNAEQVNGSINLLMRLTDDYLFIS 821
QY 868 PHLQAKFTLSTLVHGVPEYGCINLQKTVVN--FPVEPGTLGGAAPYQLPANCLPWC 925
Db 822 DSQNALNLIVOLQNCANNNGFMENDQKITNFOFPOEDYNL---EHFKISVQNECQWIG 878
QY 926 LLLDTQILEVECDYSGYAQTSIKTSLSLTSFQSVFKAGTKMRNKLKSLVLRKCHGLFDL--- 983
Db 879 KSIDMNTLEI-----KSIQKQTKQEQINQINVAISIKN-LASQKNKLSLFLNQLID 930
QY 983 ---QVANSLOTVCINIV---KIFLLQARYRHACVIQL 1012
Db 931 YFNPINISFGLCRQLYHHSKATVMPFPMTKLFQI 967

RESULT 10
O35432
ID O35432 PRELIMINARY; PRT; 67 AA.
AC O35432
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DE TELOMERASE CATALYTIC SUBUNIT (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA DRISSE R., CLEVELAND J.L.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF029235; AAB84200.1;
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 67 AA; 8368 MW; E2A06F2B CRC32;

Query Match 6.0%; Score 355; DB 11; Length 67;
Best Local Similarity 98.5%; Pred. No. 5.5e-24;
Matches 66; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 550 FFYITESTFQKNRLLFFYKRSVWSKLSQISGVHRLERLRLSELSEEVRRHQTWAMPIC 609
Db 1 FFYITESTFQKNRLLFFYKRSVWSKLSQISGVHRLERLRLSELSEEVRRHQTWAMPIC 60
QY 610 RLRFIPK 616
Db 61 RLRFIPK 67

RESULT 11
O06163
ID O06163 PRELIMINARY; PRT; 884 AA.
AC O06163
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DE CHROMOSOME XII COSMID 8543.
GN L8543.12.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.

```





```
Db 38 MDNY-----LRQNWRLTKELIKQK---YKQPVLVKEIPKPDGGRQGLGIFVMDRMIOQ 90
QY 594 EYVRHQDTWAMPICRLURFIPKPNGLRPIVNMYSMTGRALGRKQAOHQFTQRLKTLFS 653
Db 91 AIVQ-----VMSPICEPHFSDTSYGRFP--NRSC-----KAIMK 123
QY 654 MLNVERTKPHLMGSSVLGMNDIVRTWRAFLVRLALDQTPRMVFKADVGTAYDAIPQG 713
Db 134 LLEY-----LNDGYE-----WIVDIDLEKFFDTPQD 150
QY 714 KLVEVAMNIRHSSTYCIROYAVVRDSDQGVHKSFRRQVTTLSLQPYMGQLKLQD 773
Db 151 RLMSLVNHIIDGDGTESLIRKYL-----174
QY 774 SDASALRNSVVEIOSISNNESSLSLDFLFLHRLSHSVVKIGDRCYTCQCGIPIQSSSLSTL 833
Db 174 -----HSGVIINGQRYKTLVGTPOGGNLSPL 199
QY 834 LCSLCFGDMENKLPFAEVRDGL--LLRFVDDFLVTPHLDQAKTFLSLVHGVPF--YGCMI 891
Db 200 LSNI-----MLNLDLKEKRLFRVRYDCCVTVGSEAAKRVMSVSRFIEKRIGLKV 255
QY 892 NLOKTVNFPVPGTGLG 908
Db 256 NMTKTKITRPRELYLG 272
RESULT 13
O45321
ID O45321 PRELIMINARY; PRT; 561 AA.
AC O45321;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JAN-1999 (TREMBlrel. 09, Last annotation update)
DE DY3.4 PROTEIN.
GN DY3.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA LENNARD N.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., JIER M., JOHNSTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., LAISTER N., LATREILLE P.,
RA JONES M., KERSHAW J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA LIGHTNING J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA PARSONS J., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,
RA SMALDON N., SMITH A., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA THIRRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.;
RL Nature 368:32-38(1994).
DR EMBL; Z96047; CAB09413.1; -.
SQ SEQUENCE 561 AA; 65807 MW; 1C710A97 CRC32;
```

```
Query Match 2.3%; Score 133.5; DB 5; Length 561;
Best Local Similarity 19.7%; Pred. No. 0.003;
Matches 66; Conservative 65; Mismatches 149; Indels 55; Gaps 12;
QY 617 PNGLRPIV---NMSYS-MGTRALGRKQAOHQFTQRL--KTLFSMNY--ERT-KHPHLMG 667
Db 249 PHFIRPNVATFKLSLSROKRLPLFKRAIDKKTETMQKRLNSMLSLCRLSGYVIRTI 308
QY 668 SSVLGMNDIYRTWRAFLVRLALDQTPRMVFKADVGTAYDAIPQGLVEVAMNIRHSE 727
. . . . .
```

```
Db 309 DSKCKYSDF-----LKNQSNKIIIGYADVSKCFSTVNHVLYLSIIDRLFSQBH 358
QY 728 STYCIROYAVVRDSDQGVHKSFRRQVTTLSLQPYMGQLKLHLOSDASALRNSVWIEQ 787
Db 359 DIYTV-----CGKGRNHGGFHLIFCSAGTELNAHEALR--RKMEKLGVFNEVCYR 408
QY 788 STSMNESSLSLDFLFLHRLSHSVVKIGDRCYTCQCGIPIQSSSLSTLSCLCFGDMENKLF 847
Db 409 EMS-----SSTILSVIRKTTLSYTYKRGPTSMRITKGVQGHPISSNLAHMYLNNEEQYK 465
QY 848 AEVRDGLL--RFVDDFLVTPHLDQAKTFLSLVHGVPYCGMINLQKTVNFPVPEG 905
Db 466 SNEKEDSRIVFCRYEDDFITFTEKMMKPLSTGNNTNHTLTANPKK-----516
QY 906 TLGGAAPQLPAHC-----LFPWCGLLDTQTLEVF 936
Db 516 -----FKSERCASQVQLQWCGYKLDQSGNCF 543
RESULT 14
Q12815
ID Q12815 PRELIMINARY; PRT; 778 AA.
AC Q12815;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE TASTIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95278733.
RA FUJUDA M.N., SATO T., NAKAYAMA J., Klier G., MIKAMI M., AOKI D.,
RA NOZAWA S.;
RT Trophinin and tastin, a novel cell adhesion molecule complex with
RT potential involvement in embryo implantation.;
RL Genes Dev. 9:1199-1210(1995).
DR EMBL; U04810; AAA79333.1; -.
KW Cell adhesion.
SQ SEQUENCE 778 AA; 83758 MW; 4FE9DC3A CRC32;
```

```
Query Match 2.2%; Score 131; DB 4; Length 778;
Best Local Similarity 22.4%; Pred. No. 0.008;
Matches 107; Conservative 44; Mismatches 167; Indels 160; Gaps 26;
QY 2 TRAPCPAVRSLRSRYRVMVPLATFVRRLGPEGRRLVQPGDPKIYRTLVA-----QC 54
Db 358 TPMPSTPRVQ-----AQLRGVSPQS---CSEDPALPWEQVAVRLFQESC 401
QY 55 LVCMHMGSQPP---PADLSFHVSSKLKELVAR---VVORLCERNRNVLAFGELLNEAR 108
Db 402 IRSLEGGSGKPPVATPSGPHSNRTPSLOEVKIQRIQIGIQLLQRLQEVGLVGGQCVPLN---459
QY 109 GGP--PMAFTSSVRSPLNVTIETLRVSGAMLLLSRVGDDLVLVLAHALYLLVPPSC 166
Db 459 GGSLLDMVELQLLTETISRTLNATEHNSGT-----SHLPGLLKHSLG---PKPC 504
QY 167 AYQVCGSPLYQICATTDIWP-----SVSASVYRTPRPVGRNFTN-L 205
Db 505 LPECEGP--QPCPPAEPGPPEAFRCSEPEIPEPSLQEQLEVPYPAEP--RPLESCC 560
QY 206 RFLQIKSSSRQ-----APKPL-----ALPSRGTKRHLSLTSTSVPSAKK 246
Db 561 RSEPEIPESSRQEQLEVPPEPCPPAEPRLPESYCRIEPEIPESRQEQ-----EYPE---613
QY 247 ARCYPVPRVEEGPHROVLTPSGKSWVPSPARSPVEPTAEKDLSSKGKVSLSLGSV--305
Db 613 -----PCPPAEPGP---LQPSIQSGSGPPGC-----PRVELGASEPCTLEHRSLESSLPP 660
QY 305 CCKH-KPSSTSL-----SPRONAFQLRPFIEIRHFLYSRGDCQERLNPSFLLSN 354
. . . . .
```

